Affective Body Language Detection and Generation for Socially Assistive Robots

by

Derek McColl

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy

Mechanical and Industrial Engineering
University of Toronto

Affective Body Language Detection and Generation for Socially Assistive Robots

Derek McColl

Doctor of Philosophy

Mechanical and Industrial Engineering
University of Toronto

2015

Abstract

For socially assistive robots to be successfully integrated and accepted within society (especially by vulnerable populations such as the elderly), they need to be able to interpret human affective states (emotions, moods and attitudes) in order to respond appropriately during assistive human-robot interactions (HRI). Previous research has indicated that body language is a very important method of communicating affective states during natural human social interaction. This thesis focuses on developing, implementing and evaluating non-contact based autonomous affective body language recognition and classification systems that allow socially assistive robots to estimate a person's affect and respond appropriately during interactions. Both a novel static body language based recognition system and a novel dynamic body language based recognition system have been developed and implemented. The first system classifies body language using a categorical approach based on a person's accessibility (openness and rapport) towards a robot, while the second system uses a dimensional approach based on a person's valence (feeling of unpleasantness or pleasantness) and arousal (level of activity) levels. Both systems are implemented onto the human-like socially assistive robot Brian 2.1 and evaluated during one-on-one socially assistive HRI experiments. The results of experiments with young adults and elderly participants indicate that both systems accurately identify user affect.

This thesis further investigates how individuals interact and perceive an affect-aware socially assistive robot, by designing Brian 2.1 to adapt its behaviours based on a user's degree of accessibility towards it throughout social HRI. Experiments indicate that participants were more accessible towards the accessibility-aware robot over a non-accessibility-aware robot, and perceived the former robot to be more socially intelligent. With this in mind, this work lastly investigates the development of an emotionally intelligent robot capable of effectively displaying its own affective body language during such interactions. Uniquely, body language based on movements and postures founded in human behaviour research are integrated onto Brian 2.1 in order for such a robot to be able to respond to a user's affective state using its own set of emotions. The importance of these body language displays are that they are designed and tested to be easily recognizable by non-expert users. A number of social HRI studies are conducted in this work and lay a foundational framework for developing socially assistive robots that can recognize and react to human affect while providing needed assistance.

Acknowledgments

I am very grateful for the supervision of Prof. Nejat. Her guidance has been invaluable in completing my research and getting it published in international journals and conferences. I am also thankful for all the opportunities she gave me to teach, supervise undergraduates, and travel to and help at conferences. I would also like to thank my Ph.D. committee for their time and input into my research.

My thanks also goes to all the graduate students of the Autonomous Systems and Biomechatronics Laboratory for keeping the lab an enjoyable place to work.

Most importantly, I would like to thank my wife for her understanding and love during my Ph.D. Her support and encouragement is what made this thesis possible. I also appreciate the continuous support of my families, who I know are always there for me.

Table of Contents

List of Tables

List of Figures

List of Appendices

Chapter 1
Introduction

1.1 Motivation

There is currently an increase in the number of people suffering from cognitive disorders in the world. Cognitive disorders are defined as disorders that result in poor memory, and language and reasoning difficulties [1]. Examples of cognitive disorders include dementia and delirium. An estimated 44 million people worldwide are affected by dementia, this will increase to an expected 135 million people by 2050 [2].

Cognitive disorders greatly affect the quality of life of the people diagnosed and the family members who care for them. For example, an elderly person's ability to complete activities of daily living (ADLs), such as eating and hygiene, can be restricted by dementia [3], requiring the assistance of caregivers in order to effectively complete important ADL tasks. The caregiver's role covers many facets, which include providing social, cognitive and/or physical assistance. These disorders also have huge social, economic and health impacts. In 2010 alone, the healthcare costs related to dementia were as much as $604 billion worldwide [2].

There is no cure for cognitive disorders, however, it has been shown that certain pharmacological treatments and behavioural therapies can improve quality of life [4-11]. Behavioural therapy through various physical and cognitive exercises and routines has been found to be an effective method for recovery and/or minimization of cognitive disorders [5]. These types of behavioural therapies include cognitive stimulation-oriented treatments to build new ideas and associations while focusing on implicit information processing. Research studies have shown that elderly people who regularly partake in socially and cognitively stimulating activities have a lower risk of being diagnosed with dementia [6]. Social interactions have been shown to have positive effects on the physical and mental health of the elderly [7]. Similarly, cognitively stimulating interactions have been shown to be effective at improving the cognitive abilities of the elderly [8]. Additionally, cognitive assistance has also been shown to improve the ability of elderly people suffering from dementia, specifically Alzheimer's disease, to perform ADLs, namely, preparing food, face and hand washing, and getting dressed [9]. It should also be noted that cognitive stimulation and pharmacological treatments together have been shown to be more effective at increasing cognitive functioning than pharmacological treatments alone [11]. The dedicated supervision, coaching, and stimulation required for behavioural therapy are currently lacking and on a steady decline due to the decrease in health care professionals [5]. Hence, there is a need for researchers to develop robots to aid in engaging individuals in behavioural therapies.

1.1.1 Assistive Robots in Healthcare Applications

The use of assistive robots in healthcare settings is seen as one of the most important applications of robots. Assistive robots can be classified into two different groups: i) non-interactive robots for surgery, physical rehabilitation, and medication delivery, and ii) interactive robots that consist of socially assistive robots that engage in social human-robot interaction (HRI) scenarios for rehabilitation of cognitive, social and physical abilities after a stroke, accident or diagnosis of a cognitive or social disorder, and to prevent depression from isolation and deterioration of mobility [5].

1.1.2 Socially Assistive Robots

Socially interactive robots are currently being designed to engage in convincing and natural social interactions with people for a wide variety of applications. Namely, emerging applications for these robots include helpers in the home/workplace [12]; tour guides and greeters in museums, hospitals and shopping malls [13]; and aids in security and defense [14]. However, limited research has been conducted on the use and benefits of social robots as therapeutic aids or assistants for the elderly. Such socially assistive robots have the potential to bring a new interaction tool to a vulnerable population that would otherwise lack resources. They also appear to make existing care cheaper and more effective [15]. A small number of animal-like and human-like social robots have been developed to assist the elderly utilizing different combinations of communication modes including speech/sounds, gestures, facial expression, and/or a touchscreen. These communication behaviours allow the robots to assist users by providing reminders and information to complete daily tasks [16].

The seal-like robot Paro has been designed to engage elderly persons in animal therapy scenarios by learning which behaviours, consisting of sounds and movements, are desired by the way a person pets, holds, or speaks to it. Studies performed with Paro have shown that the robot can improve users’ mood and stress levels, as well as facilitate interaction between users by creating a comfortable and sociable atmosphere [17]. In [18], a human-like nurse robot, Pearl, assisted the elderly by physically guiding them to their medical appointments, providing important reminders for taking medications and providing such information as the weather and the current time. Pearl utilizes speech, facial expressions and a touchscreen to interact with users. During preliminary experiments, Pearl successfully guided elderly individuals to appointments and told them the weather and/or the time. All participants expressed a high level of excitement towards Pearl.

In [19], Fasola and Mataric have explored the role of the child-like Bandit II robot in exercise coaching. Bandit II utilizes speech and gestures to instruct elderly users through exercise routines. The results of an experiment performed at a long-term care facility indicated that Bandit II was successful at motivating elderly users to participate in exercise routines. Tapus et al., [20], also investigated the use of Bandit II at a long-term care facility as a motivator for a one-on-one music game played with elderly residents suffering from Alzheimer’s disease. During a game, Bandit II utilizes speech to give hints to the players and clapping gestures to congratulate correct responses. Experiments with elderly users showed that Bandit II was able to maintain or improve the participants’ ability to play the game over a 6 month period.

Zhang et al., [21], investigated various face, voice and interactivity characteristics for the PeopleBot robot to use for medicine delivery to the elderly. For each of the different PeopleBot configurations, the robot delivered an empty medicine bottle to a user in a simulated patient room. Both an abstract and human-like face was tested (although neither was capable of displaying facial expressions) along with a digitized human voice and a synthesized voice. The results of a post-study questionnaire focusing on perceptions of the PeopleBot robot showed that participants had increased excitement and happiness levels for the anthropomorphic robot configuration, which included the digitized human voice and the human-like face, compared to other combinations of robot characteristics.

To date, the majority of outcomes that have been utilized to study HRI with interactive robots and the elderly have focused primarily on task performance. A handful of studies have also collected detailed data on the acceptance and attitudes towards a robot and its social behaviours, however, they have been mainly focused on animal-like robots such as Paro [17], iCat [22], and Nabaztag [23].

For example, in [22], the cat-like iCat robot used speech and facial expressions as well as a touchscreen to inform an elderly user about the weather, television programs, as well as tell jokes. Their user study consisted of one-on-one interactions with the robot in a long-term care facility to investigate the level of user acceptance of iCat. Results showed that the participants' intention to use the iCat robot was significantly determined by the participants' perceived ease of use as well as their attitudes towards iCat. In another similar study of animal-like socially assistive robots, the rabbit-like Nabaztag robot [23] was utilized by elderly people as a health advisor in their homes, using speech functions to ask users about their weight and exercise routines in order to determine if they were properly following their predefined activity plan. The results of their preliminary user study showed that the elders interacted with Nabaztag, both verbally and non-verbally, and in fact, they wanted to continue using Nabaztag as a health advisor even after the completion of the study.

The majority of the socially assistive robots discussed above are either tele-operated or are limited to providing simple tasks. These robots are not capable of autonomously providing high level assistance or participating in personalized bi-directional emotional communication to keep individuals engaged in the activity at hand. In order for socially assistive robots to become effective intelligent tools in healthcare settings, there are a number of design challenges that still need to be addressed.

1.2 Challenges in Designing Socially Assistive Robots

The overall objective for socially assistive robotics is to develop intelligent systems capable of helping people recover, train, and learn [24]. Socially assistive robots can offer customizable therapy to participants, particularly those with special needs, in different settings, such as schools, healthcare facilities or personal homes. Socially assistive robots need to be socially intelligent in order to engage in natural bi-directional communication with humans to effectively provide care. Social intelligence allows a robot to share information with, relate to, and understand and interact with people in human-centered environments. Robot social intelligence can result in more effective and engaging interactions and hence, better acceptance of a robot by the intended users [25-27]. The challenge lies in developing interactive robots with the capabilities to perceive and identify complex human social behaviours, such as displays of affect and intent, and, in turn, be able to display their own behaviours using a combination of natural communication modes such as speech, facial expressions, paralanguage and body language.

Affect is a complex state of emotions, moods, and attitudes that influences a person's behaviour and cognition. Interpreting human affect is an essential part of a successful social interaction between a person and a robot. Namely, a person's affect can be used by a robot to better engage a person in an activity through its own appropriate displays of behaviour. Affect is communicated through both verbal and non-verbal communication means. Human affect models, deep rooted in psychology, will need to be utilized in order to provide accurate representation of a person's affective state. In addition, appropriate sensors and their corresponding processing mechanisms need to be developed in order for a robot to effectively identify and recognize a person's affect as provided from a number of different communication means.

In general, it has been identified that non-verbal communication, which includes body language, facial expressions and vocal intonation, convey a human's affect better than verbal expressions [28]. Body language, itself, has been found to play a vital role in conveying human intent, moods, attitudes and affect [29]. Ekman and Friesen, [30], were the first to indicate the importance of body language in conveying information concerning affective states between two individuals in communicative situations. Furthermore, a detailed review of the literature by Mehrabian showed a link between the body posture of one person and his/her attitude towards another person during a conversation [31].

Hence, in order for socially assistive robots to effectively engage users during interactions, these robots must be able to perceive and interpret a person's body language and respond with an appropriate affective display of its own. This thesis will address this challenge.

1.3 Problem Statement and Thesis Objective

The research conducted in the Autonomous Systems and Biomechatronics Lab (ASBLab) focuses on the design of autonomous and intelligent non-contact socially assistive robots for social HRI. In order for a socially assistive robot to effectively assist a person in a desired assistive activity, the robot needs to perceive and interpret the affective state of the person with whom it is currently interacting as well as respond with appropriate behaviours. In accordance with the specified assistive mission, an intelligent perception system is proposed in this thesis for use by socially assistive robots in various social and cognitive therapy scenarios. In particular, the objective of this thesis is to develop a compact real-time multimodal perception system, in both hardware and software, capable of identifying a person's affect by recognizing and classifying his/her affective body language. This information can then be used by a robot to promote HRI and its own intent through the robot's own display of appropriate emotion-based behaviours.

1.4 Proposed Methodology and Tasks

This thesis describes the development, implementation and evaluation of sensory systems that allow socially assistive robots to identify the affective state displayed by individuals' body language while engaged in HRI. Additionally this information is utilized by a socially assistive robot to adapt its own behaviours in order to effectively engage a person in the interaction. In particular, the following topics are covered in each thesis chapter.

1.4.1 Literature Review

Chapter 2 reviews the literature on the following important research areas for developing socially assistive robots capable of identifying and responding to human body language during HRI: i) human display of affective human body language, ii) socially assistive robots, and iii) the display of affective body language by social robots.

1.4.2 Detecting Affect from Static Body Language

The contribution of Chapter 3 is the development of two novel multimodal non-contact sensory systems for recognizing and classifying a person's affect from natural displays of static body poses. The first system utilizes a thermal camera and a time-of-flight camera while the second system utilizes the 2D and 3D image data from the Microsoft® Kinect™ to identify the static poses of a person engaged in one-on-one social HRI. Both sensory systems can be mounted on the robot and utilize the same technique for classifying affect from body language.

1.4.3 Detecting Affect from Dynamic Body Language

The contribution of Chapter 4 is the development of a non-contact sensory system for recognizing and classifying a person's affect from dynamic body language displays. This system utilizes learning to classify the body language of seated individuals during one-on-one assistive HRI.

1.4.4 Designing Affective Body Language for a Socially Assistive Human-Like Robot

The contribution of Chapter 5 is the design of body language displays for a human-like socially assistive robot. This chapter presents the first robotic implementation of body movements and postures described in psychology and behavioural research to correspond to specific affective states. The body movements and postures are implemented on a robot with fewer degrees of freedom than human body.

1.4.5 Socially Assistive Human-Robot Interaction Experiments

The contributions of Chapter 6 are the design of novel social HRI experiments to investigate how elderly users perceive and interact with a socially assistive robot during assistive scenarios. This chapter lays the foundation for the social HRI experiments presented in the following chapters. The HRI scenarios presented include elderly users engaged with the robot in a memory card games scenario for cognitive training and meal-eating scenario to encourage proper nutrition.

1.4.6 Detecting Affect from Static Body Language Experiments

The contributions of Chapter 7 are unique social HRI experiments to investigate the performance of the sensory systems presented in Chapter 3, which are used to identify and classify an individual's affective state from static body poses. Both systems are compared to an expert coder's ratings for balance comparisons.

1.4.7 Detecting Affect from Dynamic Body Language Experiments

The contributions of Chapter 8 are socially assistive HRI experiments to investigate the performance of the sensory system from Chapter 4, which identifies an individual's affective state from dynamic body language. Specifically, these experiments investigate the affective states communicated by the body language of elderly users engaged in an assistive HRI meal-eating scenario.

1.4.8 Affect-Aware HRI Experiments

The contributions of Chapter 9 are HRI experiments that investigate how users interact with and perceive a socially assistive robot that adapts its behaviour based on the affective state identified from their static body poses during HRI utilizing a sensory system described in Chapter 3.

1.4.9 Recognition of Affective Body Language Displayed by a Human-Like Socially Assistive Robot

The contributions of Chapter 10 are novel experiments comparing the body language displays of a human actor and a robot actor to determine how people (not trained in robotics or affective body language) recognize the body language displays developed in Chapter 5 when it is displayed by a robot instead of a human.

1.4.10 Conclusions

Finally, Chapter 11 presents concluding remarks on the developed sensory systems and HRI experiments, summarizing the contributions to the state-of-the-art and describes future possible research directions.

Chapter 2
Literature Review

In order for a socially assistive robot to respond appropriately to users during assistive interactions the robot must be capable of identifying and recognizing their affective state. As discussed in Chapter 1, body language is an important method used to communicate affect [29]. In Section 2.1, literature on both the display and recognition of human affective body language is discussed. Section 2.2 will review social robots designed to be able to identify affective human body language. Lastly, Section 2.3 will discuss robots that are able to display affective body language that can be interpreted by humans.

2.1 Display and Recognition of Human Affective Body Language

In general, body language has been categorized into four distinct classes, [32]: (i) emblems: gestures that have a direct verbal translation, (ii) illustrators: movements that are directly tied to speech, (iii) regulators: gestures that maintain and regulate a conversation, such as to tell a person to hurry up, repeat, continue etc., and (iv) adaptors: body language that conveys affective states or perform bodily actions.

Early research on adaptor style body language in [33] presented the importance of leaning, head pose and the overall openness of the body in identifying human affect. Participants were shown images of a mannequin in various body postures and asked to identify the emotion and attitude of the posture. The results indicated that posture does effectively communicate attitude and emotion, and that head and trunk poses form the basis of postural expression, with arms, hands and weight distribution being used to generate a more specific expression. More recent research presented in [34] has shown that emotions displayed through static body poses are recognized at the same frequency as emotions displayed with facial expressions. Participants viewed images of a woman displaying different poses for the emotions of happiness, fear, sadness, anger, surprise and disgust, and were asked to identify the corresponding emotions. The results showed that the body poses with the highest recognition rates were judged as accurately as facial expressions. In [35], a study performed with 60 college students utilizing stick figures showed that emotion was strongly related to varying head and spinal positions. For the study, the students were asked to choose, from a list, the emotions of 21 stick figures with three different head positions and seven different spinal positions. The emotions on the list included anger, happiness, caring, insecurity, fear, depression and self-esteem. It was found that upright postures were identified more often as positive emotions while forward leaning postures were identified more often as negative emotions. A comparison of the results with the emotional states of the participants found that the participants' own emotional states did not influence their emotional ratings of the figures. In [36], Coulson investigated the relationship between viewing angle, body posture and emotional state. Images from three different viewing angles of an animated mannequin in numerous static body poses (derived from descriptions of human postural expressions) were shown to 61 participants who identified the emotions they felt best described each image. The findings indicated that the emotions of anger, sadness and happiness were identified correctly more often than disgust, fear and surprise, and that a frontal viewing angle was the most important viewing angle for identifying emotions. It was also found that surprise and happiness were the only two emotions from the aforementioned emotions that were confused with each other. A similar study to [36] was presented in [37], where instead of an animated wooden mannequin more human-like characters were presented in images to thirty-six subjects in order for them to distinguish between postures for different expressive emotions. The subjects were asked to group the posture images into the emotions of happiness, sadness, anger, surprise, fear and disgust, and then rate the intensity of emotion expression in each image on a 5-point Likert scale. The results identified that happiness had the highest recognition rate, while disgust had the lowest. Furthermore, a different intensity level was assigned to each posture in the same emotion group.

In [38], a database of full body expressions of forty-six non-professional actors, with their faces blurred out, was presented to nineteen participants. The participants were asked to categorize the emotion displayed by the expressions based on a four alternative (anger, fear, happiness, sadness) forced-choice task. The results showed that sadness had the highest recognition rate at 97.8% and happiness had the lowest rate at 85.4%. In [39], a study was conducted to illustrate that facial expressions are strongly influenced by emotional body language. In the study, twelve participants were presented with images of people displaying fearful and angry facial expressions and body language that were either congruent or incongruent. The participants viewed the images and were asked to explicitly judge the emotion of the facial expression while viewing the full face-body combination. The results showed that recognition rates were lower and reaction times were slower for incongruent displays of emotion. Furthermore, it was found that when the face and body displayed conflicting emotional information, a person's judgment of facial expressions was biased towards the emotion expressed by the body. Comparison studies presented in [40] also investigated the influence of body expressions on the recognition of facial expressions as well as emotional tone of voice. The results reemphasized the importance of emotional body language in communication, whether displayed on its own or in combination with facial expressions and emotional voices.

Although the papers mentioned thus far focus on the display of affective states from static body poses and postures, there exists a consensus that both body movements and postures are important cues for recognizing the affective states of people when facial and vocal cues are not available [41]. In [41], point-light and full-light videos, and still images of actors using body motions to portray five emotions (anger, disgust, fear, happiness and sadness) at three levels of intensity (typical, exaggerated and very exaggerated) were presented to 36 student participants for a forced-choice emotion classification study. For the point-light videos, strips of reflective tape were placed on the actors to only highlight the motion of the main body parts including the ankles, knees, elbows, and hands, while a full-light video illuminated a person's whole body. The still images were frames extracted from the point-light and full-light videos which depicted the peak of each emotional expression. The results of the study showed that exaggeration of body movements improved recognition rates as well as produced higher emotional intensity ratings. The emotions were also identified more readily from body movements even with the point-light videos which minimized static form information.

In [42], the characteristics of a person's gait were examined to see if emotional state could be identified from walking styles. Observers examined four different people walking in an L-shaped path while displaying four emotions and then identified which emotion each walking style represented. The results showed that the emotions of sadness, anger, happiness and pride could be identified at higher than chance levels based on the amount of arm swing, stride length, heavy footedness and walking speed. In [43], the point-light technique was used to present two dances performed by four dancers (two male and two female) to 64 participants. The dances had the same number of kicks, turns, and leaps, however, had different rhythms and timing. It was found that the participants identified that certain movements corresponded to the emotions of happy and sad. Namely, the happy dance was more energetic and consisted of free and open movements, while the sad dance consisted of slow, low energy and sweeping movements. In

[44], videos of actors performing emotional situations utilizing body gestures with their faces blurred and no audio were presented to groups of young and elderly adults. A group of 41 participants (21 young adults and 20 elderly adults) were asked to label each of the videos as one of the following emotions: happy, sad, angry and neutral. A second group of 41 participants (20 young adults and 21 elderly adults) were asked to rate the following movement characteristics of the body gestures on 7-point Likert scales: 1) smoothness/jerkiness, 2) stiffness/looseness, 3) hard/soft, 4) fast/slow, 5) expanded/contracted, and 6) almost no action / a lot of action. The results with the first group showed that both the young and elderly adults were able to perform accurate emotion identification, however, the elderly adults had more overall error especially with respect to the negative emotions. With respect to movement characteristics, it was found that the angry body language was identified to have the jerkiest movements, followed by happy, while sad and neutral had the smoothest movements. In addition, angry was rated to have the stiffest movements followed by sad. Happy and neutral had the least stiff movements. Lastly, the body movements for happy and angry were found to be faster and have more action than those for sad and neutral. In [45], arm movements performing knocking and drinking actions which portray the ten affective states of afraid, angry, excited, happy, neutral, relaxed, strong, tired, sad and weak were presented as point-light animations to participants. Fourteen participants were asked to categorize each point-light animation as one of the aforementioned ten affective states. It was found that the level of activation of an affective state was more accurately recognized for the arm movements than pleasantness using a two-dimensional scale similar to the circumplex model [46].

In [47], Wallbott investigated the relationship between body movements and postures, and fundamental and social emotions. The movements and postures included collapsed/erected body postures, lifting of the shoulders, and head and arm/hand movements. Six female and six male professional actors performed 14 different emotions. Twelve drama students acted as expert coders to identify a sample of videos which had the most natural and recognizable emotions of the actors. Then these videos were coded by two trained observers. The 14 emotions considered were elated joy, happiness, sadness, despair, fear, terror, cold anger, hot anger, disgust, contempt, shame, guilt, pride and boredom. Inter-observer agreements of 75%-99% were found for the body movement categories representing the upper body, shoulders, head, arms and hands. Wallbott found that statistically significant relationships exist between specific movements and postures of the body, head and arms, and each of the 14 different emotions. For example, boredom can be characterized by a collapsed upper body, an upward tilted head, inexpansive movements, low movement activity and low movement dynamics. The results of the discriminant analysis resulted in a 54% correct classification for all the emotions with shame having the highest correct classifications at 81%, followed by elated joy at 69%, hot anger at 67% and despair, terror and pride with the lowest classification percentages at 38%.

In [48], de Meijer investigated the relationship between gross body movements and distinct emotions. The body movements studied included trunk and arm movements, movement force, velocity, directness, and overall sagittal and vertical movements. Eighty-five adult subjects were shown ninety-six videos of three actors performing these various body movements and asked to rate the compatibility of the body movements, on a 4-point Likert scale, with respect to 12 emotions: interest, joy, sympathy, admiration, surprise, fear, grief, shame, anger, antipathy, contempt and disgust. The results showed that the participants rated the majority of the body movements as expressing at least one emotion. Furthermore, it was determined that a unique combination of body movements was utilized to predict each distinct emotion. For example, a stretching trunk movement while opening and raising the arms would lead to the subjects selecting the emotion joy.

The aforementioned literature review has shown the importance of body language in communicating human affect. In particular, it has been determined that specific body poses, postures and movements can be directly recognized as affective states. Therefore, in order to achieve effective social HRI, it is important for a socially interactive robot to be able to recognize a person's display of affective body language.

2.1.1 Automated Human Body Language Identification and Classification

There has been a lot of previous work utilizing various sensors and approaches for identification of human bodies, i.e., [49, 50], or body parts and their respective poses, i.e., [51-71]. In addition to the literature review that is presented here, there are also comprehensive literature surveys presented in [72-74]. The literature review presented here provides a discussion on body language identification approaches that have relied on the use of non-contact sensory systems. In general, a number of techniques have focused on utilizing non-contact sensing means via either a single camera [51-56] or multiple cameras [57-62].

2.1.1.1 Single Camera Techniques

In [51], Juang et al. utilized a single color camera to perform human body posture estimation using body silhouette and skin color information. First, the body is segmented from the background using a background difference technique and then, posture estimation is performed using shape features and skin color information to locate the head, the tips of the feet and hands. In [52], Mori et al. utilized the Normalized Cuts algorithm in order to group similar pixels in a 2D color image into regions known as superpixels to identify the limbs and torsos of people. The configuration of the body parts were determined by applying limb and torso detectors containing global constraints such as relative widths, lengths, location and symmetry in clothing. In [53], Chen et al. proposed the use of a deformable template matching framework which uses geometric blur descriptors and integer quadratic programming to estimate human pose using exemplars in a 2D color image of a person's profile. Experiments verify the ability of the technique to identify limbs and torso of humans walking on a treadmill. Alternatively in [54], a thermal camera is used to identify individuals that are lying down within a dense group of people. A head detection method based on local peak, elliptical shape and head-shoulder pattern detection is used to first identify people in a thermal image. A height threshold is used to label a person as either standing or lying down.

Researchers have recently also considered the use of Time-of-Flight (TOF) cameras to estimate 3D human body information [55, 56]. In [55], human arm gestures were determined by applying a 3D double difference image to extract a person's arm (the only moving body part) and represent the arm using harmonic shape contexts. A correlation-based matching technique was used to determine the different gestures such as point right, raise arm, clap and wave. In [56], a driver's pose was determined using the depth and intensity images provided by a TOF camera. In order to estimate body pose, an articulated iterative closest point (ICP) algorithm was used to register the 3D data points provided by the sensor to a 3D articulated body model. The articulated body parts include the limbs, head and torso. User specified joint constraints were used to limit possible body poses for the proposed application.

2.1.1.2 Multiple Camera Techniques

The body pose estimation techniques that utilize multiple cameras have mainly consisted of a number of 2D cameras to provide multiple view images, i.e., [57-59]. In [57], Kohli et al. used multiple 2D cameras to perform segmentation and 3D pose estimation of a human body. Their technique was based on optimizing a cost function using a Conditional Random Field approach. This cost function was optimized over all pose parameters using dynamic graph cuts to provide both segmentation and body pose. In [58], 2D evidence from images provided by multiple cameras placed in an environment was combined using epipolar geometry to obtain high likelihood body part regions in 3D. Regions with high priors for a given part were obtained using the probability distribution of connecting parts. The proposed approach was capable of addressing occluded body parts. In [59], Van den Bergh et al. use multiple 2D cameras to perform 3D pose estimation of a human body for a virtual reality application. The silhouettes of a person are formulated from multiple static cameras placed in the environment. A 3D voxel representation of the pose is formulated from the silhouettes and compared to a library of 3D haarlets that represent a database of known poses. Recently, other approaches consisting of a combination of 2D and thermal cameras [60], or a TOF camera and a 2D camera [61] have also been proposed. In [60], multi-camera views provided by four 2D cameras and four thermal cameras were used to track a driver's head, gaze direction and hands while driving. Thermal thresholds were used to segment the head and hands for 3D tracking by triangulation. 2D color information was used to determine head position by obtaining projection profiles and head orientation using a moment-based classifier. Knoop et al in [61] used a TOF camera and a 2D color camera to identify and track the pose of a human by matching 3D data to a predefined body model of the human utilizing an ICP algorithm.

2.1.1.3 Automated Affective Body Language Classification Techniques

To date, a great deal of work has been conducted on developing automated techniques utilized in determining affective states through facial expressions including Bayesian networks [75, 76], optical flow techniques [77], artificial neural networks [78], and fuzzy logic [79]. An overview of proposed techniques can be found in [80]. Only recently, a handful of researchers have also focused on recognizing a person's affective state via human body language [81-102]. In comparison with facial expressions, body language recognition provides a more complex problem due to the increased degrees-of-freedom and significant changes in the shape of the overall body [83].

In [84], hand movements of a person are identified and tracked in 2D images in real-time by using moving skin masks defined by skin color information. These movements are identified automatically to represent hand gesture classes such as clapping, hands over the head, lifting of the hands, and Italianate hand gestures using a hidden Markov model (HMM) classifier. The hand gestures are defined to represent six human emotions. These hand gestures are only utilized to support the outcome of the main facial expression analysis system. Similarly, in [85], body gestures are identified through tracking of a person's hands and shoulders in 2D images and used as supplement information to a facial expression system to identify a person's emotions. The gestures are classified into labeled emotions, using the BayesNet Weka tool for automated classification, after the interaction has taken place. In [86], an automated system detecting facial actions in videos has been developed. Head and shoulder movements are tracked using cylindrical head tracking and auxiliary particle filtering to increase the accuracy of determining a posed or spontaneous smile via a Gentle Support Vector Machine classifier. Schindler et al. in [83] used 2D still images of actors to extract emotions from their body pose using a neural model. Body language was related to an emotion category using a support vector machine classifier. In [87], manually labeled unintentional hand gestures of students during classroom lectures were used to predict their mental states using a dynamic Bayesian network-based computational model. These mental states included stressed, tired, thinking, satisfied, recalling, and concentrating. Kapoor and Picard have developed a real-time multi-sensor system using a camera to detect facial expressions and a pressure sensing chair to determine postural shifts of children trying to solve an educational puzzle on the computer [88]. Both sensory inputs as well as specific features of the activity (e.g. level of difficulty and game state) are utilized to determine affective states of interest and boredom via Gaussian process classification. Castellano et al. in [89] investigated the relation between non-verbal behaviour and emotion of a child playing chess with a robot. The video of the interaction was segmented manually after the interaction to identify non-verbal behaviour. Castellano found that certain behaviours, for example looking at the robot, gave a good indication that the child was experiencing a positive feeling. In [90], De Silva et al. proposed a real-time gesture recognition system to determine and estimate the intensity of a child's emotions during game playing situations. The gesture descriptions module obtains body part motion utilizing a motion capture system with eight markers placed on a child's trunk, lower arms, upper arms and head. The motion information is then utilized as an input into the gesture recognition model based on HMMs, so extensive training of the system is required. The system requires contact-based sensing to determine the emotions of the children. Castellano et al. presented an emotion recognition system based on gesture dynamics captured from video [91]. In particular, they explore the use of both a dynamic time warping approach and feature-vector-based classifiers to classify full body gestures of ten participants acting out anger, joy, pleasure and sad emotional states in a controlled setting. They propose that more classification schemes and better model selections will need to be explored in order to develop a real-time system capable of determining real-life emotions.

In [92], Gong et al. utilized shape of Gaussian descriptors to identify feature vectors of a person's movements during a knocking action with and without personal biases, where a personal bias was defined as the mean of an individual's feature. Classification was tested on the biased and unbiased motions with a support vector machine (SVM) classifier. Experimental results showed that the affective states of the biased motions were recognized at a rate of 59% while the unbiased motions were recognized at a rate of 76%. In [93], SVM classifiers were utilized to recognize affective states from sequences of acted actions presented in [103] in the following order: walking, lifting, walking, knocking, walking, and throwing. The feature vectors utilized for affective state classification included the means and standard deviations of joint positions, speeds, accelerations and jerk. A comparison was performed between isolated action models and connected sequential action models. Experimental results indicated an increase in the recognition rates from 52% (isolated action models) to 81% (connected sequential action models). In [96], five dancers performed the exact same dance with four different affective states: anger, fear, grief and joy. Body language features, obtained from 2D image analysis, included quantity of motion, contraction index, upwards movement, length and direction of overall body motion, and velocity and acceleration of body parts. Using a decision tree classifier, an average recognition rate of 36% was achieved for the four affective states.

In [99], affective body poses displayed by people while playing a video game obtained from a worn exoskeleton motion capture system were automatically identified using a multi-layer perceptron (MLP) approach. Body pose features that were utilized for recognition included the Euler angles of the head, neck, collar, shoulders, elbows, wrists, torso, hips and knees. With respect to these features, the MLP achieved a recognition rate of 66.7% for identifying the affective states of concentrating, defeated, frustrated, and triumphant. With respect to valence and arousal dimensions, the MLP had recognition rates of 84.2% and 82.9%, respectively. In [100], dynamic body movements, recorded during the same experiments presented in [99], were utilized to recognize the affective states of high intensity negative emotion, happy, concentrating, and low intensity emotion using a recurrent neural network (RNN). The body movements included the angular velocity, acceleration and frequency of the right arm and hand, as well as the directionality of the spine and head. The RNN obtained recognition rates ranging from 36%-67% for the aforementioned affective states.

In [101], the Intelligent Tutoring system was used to identify if twenty-four middle school students were either frustrated or not frustrated while completing a Towers of Hanoi virtual puzzle. The students were seated in a pressure sensitive chair that detected body posture. Additional sensors were also used to detect head and hand gestures, facial expressions and skin conductance. These features were utilized as input into different machine learning techniques to classify frustration. The Gaussian process classification and SVM (with a Gaussian process kernel function) techniques both obtained the highest recognition rate of 79.2%.

2.2 Automated Affective Human Body Language Recognition for Robots

A number of body language recognition systems for robots have been proposed in the literature, but have primarily focused on face, hand or arm recognition for input commands to a robot in order for it to accomplish a certain task [63-71]. For example, in [63], stereo cameras and an arm model fitting algorithm were used to determine human arm poses that can potentially be used to control a robot. In [68], an arm gesture recognition technique was developed to control a mobile robot during a clean-up task. Images from a 2D camera were used to recognize the pose of the arm using either a pose template matcher or a neural network based method. In [70], Burger et al. utilized video streams from stereo cameras to locate a person’s head and hands via an interactively distributed multiple object tracking particle filter. 3D poses of the head and hands were achieved using a sphere (for the head) and ellipsoid (for the hands) fitting technique. In [71], the detection of humans in cluttered scenes for HRI applications using a stereovision system and a thermal camera was presented. Namely, a depth-oriented scale-adaptive filter and head-shoulder contour technique were used to identify human-like objects in a scene utilizing stereo information, and a thermal image was used to identify human thermal features such as the face or hair for human verification. This technique, however, does not identify the pose of the human body or body parts.

The only system that has identified affective states from body language during HRI was presented in [102]. Utilizing a 2D environmental camera, seated children were detected to determine if they were engaged or not engaged in a chess game with the iCat robot. The following body posture features were identified from the children: body lean angle, slouch factor, quantity of motion, contraction index and meta features consisting of the derivatives of the aforementioned features over time. Testing the system with various machine learning techniques showed that the ADTree and OneR techniques obtained the highest recognition rate of 82%.

2.3 Robots that Display Affective Body Language

A number of robots have been designed to display specific emotions through dance, i.e., [104-110]. In particular, some researchers have utilized Laban body movement features from dance to generate robot emotions, i.e. [104-108]. Laban movement analysis investigates the correlation between a person's body movements and his/her psychological condition [111]. For example, a movement that is strong, flexible and has a long duration gives a psychosomatic feeling of relaxation. The four major Laban movement features are defined as space, time, weight and flow [111]. Space relates to whole body movements, it measures how direct, open and flexible the body movements are. The time feature determines the speed at which body movements travel spatially, i.e., if a body movement is sudden or sustained. Weight determines the energy associated with movements, i.e., if they are firm or gentle. The flow feature is concerned with the degree of liberation of movements, identifying if movements are free or bound. In [104], Laban features were utilized to create dancing motions for a mobile robot with 1 rotational degree of freedom (DOF) for each arm (two arms in total) and 1 DOF for head nodding. The robot performed six different dances, each displaying one of the following emotions: joy, surprised, sad, angry or no emotion. In [105] and [106], Laban dance features were used to define the motions of the small 17 DOFs KHR-2HV human-like robot for the emotions of pleasure, anger, sadness and relaxation. In particular, in [105], each of these emotions was attributed to only three distinct body movements which consisted of raising and lowering the arms. In [107], Laban dance theory was utilized to describe the body movements of a teddy bear robot. Arm and head motions of the robot were attributed with the emotions of joy, sadness, surprise, anger, fear and disgust. In [108], the 17 DOFs small humanoid Robovie-X robot generated dance movements to express the emotions of anger, sadness and pleasure based on Laban movement analysis and modern dance using its upper body, head, arms, hands, legs and feet.

Other robots have also been designed to mimic human emotional dance without utilizing Laban movement features, e.g. [109, 110]. For example, in [109], the Sony QRIO robot was used to imitate the dance motions of a person in real-time using moving region information, with the goal to create sympathy between a person and the robot. In [110], the Expressive Mobile Robot generated emotionally expressive body movements based on classical ballet using 7 DOFs in its arms, head and wheels. Experiments were conducted to see which body movements people found natural as well as which body movements depicted a feeling of interest by the robot.

A relatively small number of robots have also been developed to display emotions using body movements without incorporating emotional dance. For example, Keepon, a tele-operated chick-like robot utilizes the body movements consisting of bobbing, shaking, and swaying to convey the emotions of excitement, fear and pleasure, respectively [112]. The robot has been designed for interactions with children diagnosed with autistic spectrum disorders. In [113], the design of an insect-like robotic head with two arm-like antennas was presented to express different emotions using exaggerated expressions of animated characters. Namely, the change in color of the eyes and antennas, the motion of the antennas and the eye emoticons can be used to display such emotions as anger, fear, surprise, sadness, joy and disgust. Examples for expressive antenna motions include the ends of the antennas being brought in front of the eyes for fear and swept backwards for surprise. In [114], the small humanoid robot Nao was utilized to express the emotions of anger, sadness, fear, pride, happiness and excitement through head movements in a range of different robot poses. The poses of the robot were designed based on motion capture information of a professional actor guided by a director. In [115], the human-like WE-4RII robot was used to display emotions using facial expressions and upper body movements (especially hand movements). The facial and body patterns to display for the emotions were based on recognition rates from a pre-experiment where several simulated patterns were presented to subjects. Both the posture and velocity of the body were used to display the emotions of neutral, disgust, fear, sadness, happiness, surprise and anger. In [116], the Nao robot was also utilized to generate the emotions of anger, fear, sadness and joy with body movements, sounds (i.e., crying, growling, banging), and eye colors (i.e., red for angry, dark green for fear, violet for sad, yellow for joy) in order to map these emotions onto the Pleasure-Arousal-Dominance (PAD) model. The authors stated that they used psychological research inspired by the work of Coulson [36] and de Meijer [48], TV shows and movies to link emotions to body movement, sound and color. Expressions did also include dancing for the emotion of joy and saying “Jippie Yay” with the robot’s eyes turning yellow. The robot’s emotion expressions were first evaluated in a pre-test and then each single expressional cue was individually investigated in the experiments in order to determine the expressivity of each stimulus for each emotional cue. However, for these expressions, the authors did not specify which descriptors from Coulson and de Meijer they considered and for which emotions. Hence, it is not clear how the poses/movements of the small robot are directly linked to existing human psychological studies.

In general, the emotions of robots designed for HRI have mainly been derived from body movements from dance or robot-specific characteristics. For the latter group, robot-specific movements have usually been generated that cannot easily be generalized to other robots. With respect to emotional dance, the corresponding body movements are more appropriate for small robots that can have a larger workspace (i.e., table tops) during HRI, and cannot be effectively used for larger robots engaging in natural one-on-one social interactions. To date, research into the use of emotions based on human body movements and postures for social interactions is non-existent for robotic applications with the exception of [116]. However, in [116], emotional dance is still incorporated into some of the small sized Nao’s emotional expressions and the link between the robot’s body language and human body language is not directly clear. Hence, this research explores the challenge of using natural human body movements and postures to represent social emotional behaviours for life-sized *human-like* robots in order for the robots to effectively communicate while building interpersonal relationships during one-on-one social interactions.

2.3.1 Human Perception of Robotic Affective Body Language

A handful of researchers have also investigated human perception of robot body language in representing specific affective states. In [114], the head positions of Nao were utilized to investigate the creation of an affect space for body language. Twenty-six participants were asked to identify the emotions displayed by the robot, based on different head movements, as anger, sadness, fear, pride happiness or excitement. Participants were also asked to rate the level of valence and arousal of each emotion utilizing a 10-point Likert scale. The results showed that a head-up position increased the recognition rates of the emotions of pride, happiness and excitement, and a head-down position increased the recognition rates of the emotions of anger and sadness. The position of the head was also found to be related to the perceived valence of the robot's emotion but not to arousal. In [115], the human-like robot WE-4RII was utilized to determine how well participants could recognize the emotions of the robot utilizing facial expressions, body and hand movements. It was found that the participants recognized emotions more often when emotional hand movements were included with facial expressions and body movements. In [116], sixty-seven participants were asked to identify which combination of body movements, sounds, and eye colors that the Nao robot displayed were most appropriate for the emotions of anger, fear, sadness and joy. Then another study was conducted with forty-two participants, where the robot separately displayed body movements, sounds and eye colors for the same emotions. In this latter study, the participants were asked to assign a specific value within the PAD model for each of the individual expressions. It was found that body movements achieved the best results. In [117], one set of participants (which included amateurs and expert puppeteers) was asked to create simple non-articulated arm and head movements of a teddy bear robot for different scenarios. Another set of participants was asked to watch animations or videos of these robotic gestures and to judge the emotions that were displayed based on the simple movements created. The emotions that were available for the second set of participants to choose from included happy, interest, love, confused, embarrassed, sad, awkward, angry, surprised and neutral. The participants also rated the lifelikeness of the gestures and how much they liked the gestures. The results showed that emotions can be conveyed through simple head and arm movements for the teddy bear robot and that recognition rates increased when the participants were given the situational context for the gestures. The gestures for fear and disgust were found to be better understood when created by expert puppeteers rather than amateurs, however, this was not true for the other emotional movements. It was also found that positive emotions and more complex arm movements were rated as more lifelike.

Studies determining recognition rates of emotions based on the use of Laban body movements have also been conducted. For example, in [108], emotional dance for the three basic emotions of anger, sadness and pleasure was displayed by the small humanoid Robovie-X robot to two different groups of Japanese participants. In particular, a group of elderly individuals and a group of young individuals were asked to watch and identify each emotion displayed by the robot's body movements. The results showed differences in the perception of emotion from robot body language between the two groups. The authors suggested that these differences are due to variations in the focus and cognition of the two groups when identifying the emotions such as their attention to different body parts and their perception of the magnitude and speed of the robot's motions. Hence, body language of the robot should be designed with the consumer in mind. In [105], thirty-three subjects watched the KHR-2HV human-like robot's movements and categorized these movements as being a weak or strong display of pleasure, anger, sadness or relaxation. The subjects first watched the robot display basic movements and then eight processed whole-body movements which represented the target emotions. The results showed that the subjects could identify the emotions of sadness, pleasure and anger for the movements but not relaxation, and that some emotions could easily be confused with each other such as pleasure with anger, and sadness with relaxation. In [107], eighty-eight Japanese subjects were asked to identify the emotions related to the Laban body movements displayed by a teddy bear robot with 6 DOFs in the head and shoulders. The emotions were chosen from a list which included joy, anger, surprise, fear, disgust and sadness. They were also asked to rate on a 4-point Likert scale how clearly the emotions were displayed. The results found that with simple arm and head movements, the emotions of joy, sadness, surprise and fear could be recognized. However, anger and disgust were not easily recognized by the subjects. In [104], twenty-one student participants were asked to judge the intensity and type of emotions (joy, surprised, sad, angry and no emotion) displayed by a mobile robot to determine correlations between these emotions and the robot's effort and shape movement characteristics that are based on Laban movement features. Effort represents dynamic features of movement or quality of movement, whereas shape represents geometrical features of the overall body. The results showed that strong body movements were correlated with joy, and they were also correlated along with ascending and enclosing shape features to surprise. Weak body movements were correlated with sadness and an advancing body movement was correlated with angry.

None of the robotic studies presented above compared the recognition rates of emotional body language displayed by a robot with the recognition rates of the same emotional body language displayed by a human actor. Such a comparison study allows for the investigation into the quality of the body language displayed by the robot. This will allow for the determination of which body movements and postures can be generalized for the robot to display for a desired emotion, in addition to exploring whether human body language can be directly mapped onto an embodied human-like robot.

Chapter 3
Detecting Affect from Static Body Language

In this chapter, a novel automated affect from static body language recognition system is developed for social robotic applications capable of interpreting, classifying and responding to natural body poses during HRI. The overall architecture of the proposed methodology is shown in Figure 1. 3D sensory information of a person as well as skin region information of the lower arms and head as obtained from the viewpoint of the robot are used as inputs into the methodology. Body part segmentation is performed using both the 3D and skin region information. Static body language of a person is identified by tracking these individual body parts during an interaction. A 3D ellipsoid model is utilized to estimate the 3D poses of the body parts during the display of body language. Once a 3D body pose has been identified, it is used to classify the person's affect. This system identifies affect as a person's degree of accessibility towards a social robot. A person's degree of accessibility refers to his/her psychological state of openness and rapport towards another during dyadic social interactions [118]. Previous research has found a significant relationship between an individual's accessibility and his/her body pose [119]. This chapter presents an automated accessibility recognition system that utilizes and adapts the position accessibility scale of the Nonverbal Interaction and States Analysis (NISA) [118, 119] to identify an individual's degree of accessibility utilizing his/her trunk and arm orientations towards 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 [119, 120]. This work extends the use of NISA to social human-robot interactions. Prior to providing a detailed explanation of the proposed body language recognition and identification system, the body poses of interest are described.

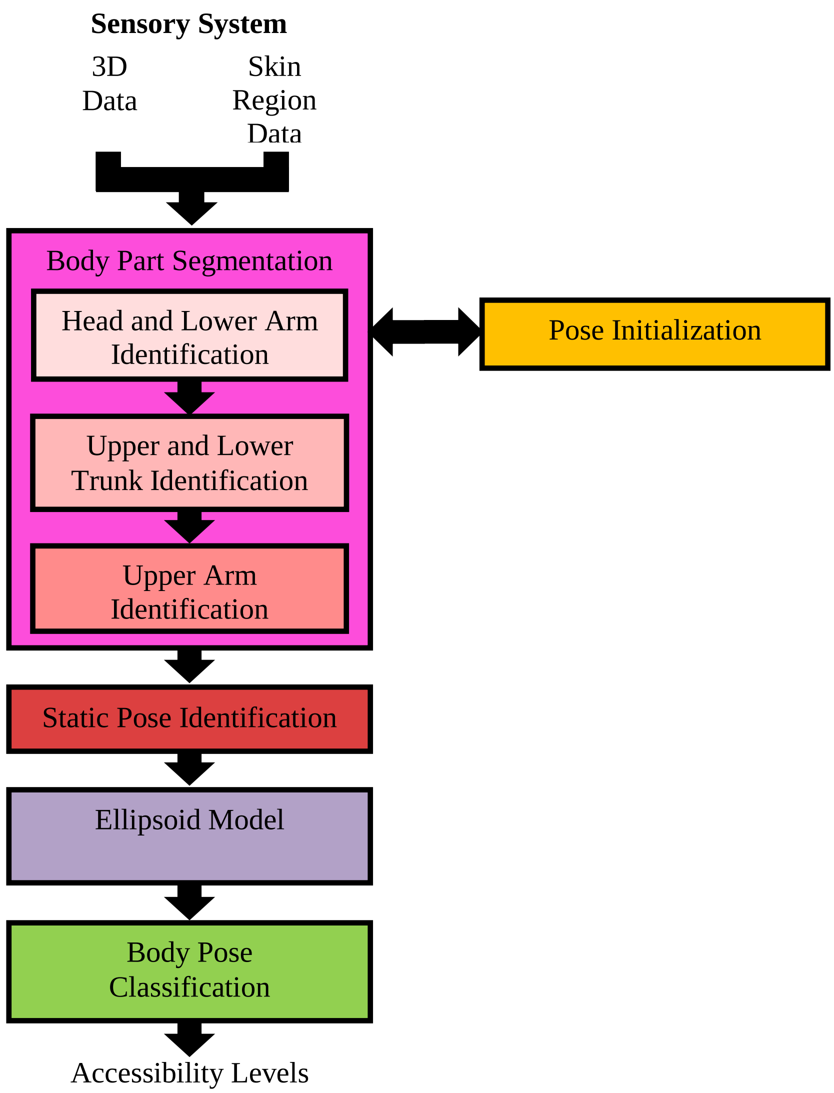


Figure 1 Static Body Language Classification Architecture. [121]

3.1 Defined Static Body Poses

While engaged in social interaction scenarios a person can display different static body poses. Herein, a body pose is defined to be static if it is held for at least four seconds [118]. The static body positions defined in the position accessibility scale are an arrangement of trunk orientations and leans, and arm positions which are utilized to determine the accessibility of a person with respect to the robot, as summarized in Table 1. In particular, the orientations of the upper and lower body trunks are coded as Toward (T) - when the person's trunk is oriented at an angle from 0^o to 3^o from the robot; Neutral (N) - when the person's trunk is oriented away from the robot by 3^o to 15^o; and Away (A) - when the person's trunk is angled more than 15^o from the robot. Trunk lean is described as upright - when the person's shoulders are directly above the hips; and forward - when the person's shoulders are shifted in front of the hips A finer-grained scale can also be applied using different arm arrangements: (T) - where the person's arms are positioned in a such a way that has them nearer to the robot than the person's upper trunk, (A) - where the arms are positioned in such a way that has them further away to the robot than the person's upper trunk, and (N) - where the arms move symmetrically with the upper trunk and are not positioned independently.

Table 1 Static Body Pose Orientations and Leans.

Trunk Orientation	Trunk Angle with Respect to Robot
Away (A)	$> 15^o$
Neutral (N)	$3^o – 15^o$
Toward (T)	$0^o – 3^o$
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

3.2 Human Body Extraction

Since social robots can be utilized in a large variety of 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 3D data of the person interacting with the robot from the scene, a technique was developed that utilizes a combination of Mixture of Gaussians (MOG) [122], connected component analysis [123], and head and shoulders contours [124]. A statistical model of the environment is generated by creating an MOG for each pixel of a 3D depth image utilizing multiple training depth images of the scene (without people), prior to the interaction scenario between the person and the robot. During the one-on-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. Connected components that are able to be fit with a head and shoulders contour are classified as a person. Finally, the person that 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, the background noise can also be removed from other imaging sensors by calibrating to the 3D depth sensor, isolating only the user in the images of the other sensors.

3.3 Pose Initialization

An initialization stage is utilized to aid segmentation of the upper and lower trunks of the person during HRI. During the initialization stage, the person stands in a neutral position with his/her arms alongside the trunk with no occluded body parts. Anthropometric measurements are utilized to automatically identify the location of the person's waist and hips in this neutral pose. It is assumed that during the interaction, the heights of a person's waist and hips do not significantly change as the robot maintains a constant distance from the person. The anthropometric information is determined within the sensory information. In general, it has been found that for the 50th percentile of males and females, the height of the waist of a person standing in a neutral pose is within one standard deviation of the height of the elbow, and the height of the hips is within one standard deviation of the halfway point between the height of the elbow and the knuckles [125, 126]. This relationship is utilized to approximate the heights of the waist and hips based on the skin information of the lower arms. Namely, within the skin information, the regions representing the lower arms are identified based on the location of their centroids. Bounding boxes are drawn around the lower arms and then used to approximate the locations of the waist and hips based on the average heights of the upper and lower edges of the boxes.

3.4 Head and Lower Arm Identification and Segmentation

Skin region information is utilized to identify the location of the head and lower arms, as these are the body parts that are/can be 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 [66, 70, 127, 128]. This information can be obtained from different imaging sensors such as 2D cameras or thermal cameras placed on the robot. A detailed discussion on the integration of the proposed methodology with different sensory systems is presented in Subsection 0 of this chapter. Skin region information of the lower arms and head is represented using binary images, examples of which can be seen in Figure 2. In general, the number of skin regions identified can vary between 1 and 3 based on the NISA body poses displayed by a person. Once these skin regions are identified, the location and pose of the head and lower arms can be estimated.

(a) (b) (c)

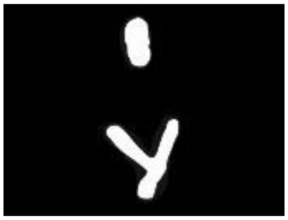
Figure 2 (a) one, (b) two, and (c) three skin regions.[129]

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 and head touching. 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 2. 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, [130], was utilized to determine the most appropriate machine learning technique to utilize for classifying head and/or lower arm configurations. A 10-fold 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. Multi-Layer Perceptron); and 7) non-linear 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, [131], had the highest recognition rate of 99.3% and has been implemented in the system architecture. Once all the skin regions have been classified, the regions containing multiple body parts, e.g. Figure 3, are further segmented to identify individual parts.

Table 2 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 R is the total number of skin regions, and N_i is the number of pixels in the 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 a and b are the lengths of the major and minor axes of an ellipse fitted to the i^{th} skin region.	(4)
Expansiveness	Normalized (x,y) pixel coordinates of two opposite vertices, $V_{i,1}$ and $V_{i,2}$, of the skin region's axis-aligned bounding box.	$V_{i,1} = \left(\frac{\min_j (x_{i,j})}{T_x}, \frac{\min_j (y_{i,j})}{T_y} \right)$,	(5)
		$V_{i,2} = \left(\frac{\max_j (x_{i,j})}{T_x}, \frac{\max_j (y_{i,j})}{T_y} \right)$	(6)


(a) Two arms touching


(b) Arm clasping neck

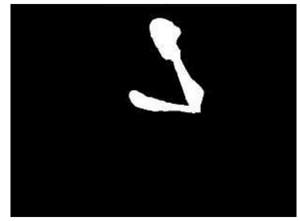
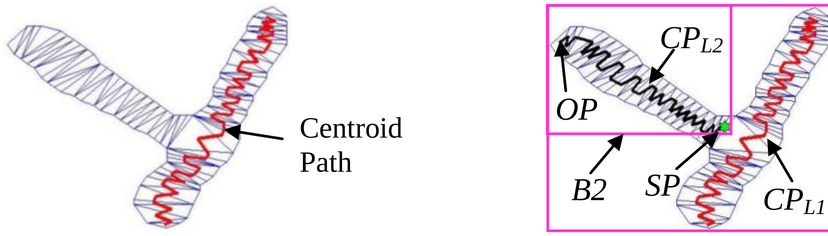

(c) Arm touching head and other arm

Figure 3 Examples of poses with touching body parts. [129]

A technique that utilizes the Delaunay triangulation was developed to segment two touching body parts from each other. In particular, Delaunay triangulation is performed on the contour of the region encompassing the two touching body parts. Figure 4 depicts the triangles determined by Delaunay triangulation for the pose in Figure 3(a). The centroid of each triangle is formulated and paths are formed along the centroids of adjacent triangles. In general, there can be no overlapping paths. The two centroid paths with the largest number of centroids, CP_{L1} and CP_{L2}, are identified to determine where to segment the two touching body parts. The distances from the endpoints of CP_{L2} to every centroid along CP_{L1} are calculated, and the endpoint of CP_{L2} which has the minimum distance to a centroid on CP_{L1} is identified to be the location at which the two body parts are to be separated. This centroid is labeled as the separation point *SP* in Figure 4(b). The remaining endpoint of CP_{L2} is labeled as *OP*. A bounding box, *B1*, is then drawn around the region that contains the two touching body parts. A second box, B2, is drawn from SP to the corner of B1 that is closest to OP. An example of this is shown in Figure 4(b). The region inside B2 is defined as one body part and the remaining region of B1 not including B2 is defined to be the other body part. To determine if each of the separated body parts is a head or an arm, the body parts are tested with the Adaboost learning technique. The identified bounding boxes for the two poses in Figure 3 are depicted in Figure 5 for both a two arms touching and a one arm touching the head configuration.

If a skin region is classified as both lower arms and head touching, e.g., Figure 3(c), segmentation of the individual body parts is achieved by implementing two iterations of the Delaunay triangulation procedure used to segment two body parts. First one body part is segmented from the skin region and then the two remaining body parts are segmented from each other, the resulting bounding boxes for each segmentation iteration of the pose in Figure 3(c) are shown in Figure 6.



(a) Delaunay triangles and centroid path (b) Intersection point between body parts

Figure 4 Triangulation of contour for region consisting of two arms touching. [129]

(a) Separation boxes for image in Figure 3(a) (b) Separation boxes for image in Figure 3(b)

Figure 5 Example segmentation results of two body parts touching. [129]

(a) 1st iteration of segmentation (b) 2nd iteration of segmentation after first segmented body part was removed

Figure 6 Example segmentation results for three body parts touching.

Table 3 represents a comprehensive list of the different potential head and lower arms configurations that the automated region-based segmentation technique can address. The configurations presented herein are based on the natural static poses that can be classified by the position accessibility scale of NISA. It can be noted that additional body poses can exist in which the number of skin regions can exceed three. However, this can only happen when a nonexposed body part segments a lower arm or head into at least two regions. However, these poses are rare, and they do not reflect natural body language.

Table 3 Detectable Head and Lower Arm Configurations.

Number of Skin Regions	Head and Lower Arm Configurations
3	• Head and both lower arms generate distinct high intensity thermal regions
2	• Head generates one high intensity thermal region, however, the lower arms are touching each other and together generate another high intensity thermal region • One lower arm is touching the head creating one high intensity thermal region, and the other lower arm generates a distinct high intensity thermal region • Head and one lower arm generate distinct high intensity thermal regions, and the other lower arm is occluded • Both lower arms generate distinct high intensity thermal regions, and head is occluded.
1	• Both lower arms are touching the head generating one high intensity thermal region • Both lower arms are touching each other generating one high intensity thermal region, and head is occluded • One lower arm is touching the head generating one high intensity thermal region, other arm is occluded • One lower arm generates one high intensity thermal region, head and the other lower arm are occluded • Head generates one distinct high intensity thermal region, and both lower arms are occluded

3.5 Upper and Lower Trunk and Upper Arm Identification

Once the location of the head and lower arms are determined using skin region information, this information can be superimposed on the 3D data of the person in order to determine the remaining body parts. The locations of the waist and hips formulated during initialization, and the location of the shoulders determined as the minimum height of the region defined as the head, are utilized to identify the upper and lower trunks in the 3D information. The upper trunk is defined from the shoulders to the waist and the lower trunk is defined from the waist to the hips, as seen in Figure 7. The width of both the upper and lower trunks are determined by finding the width of the 3D data that represents the person at their waist height with the 3D data of the head and lower arms removed.

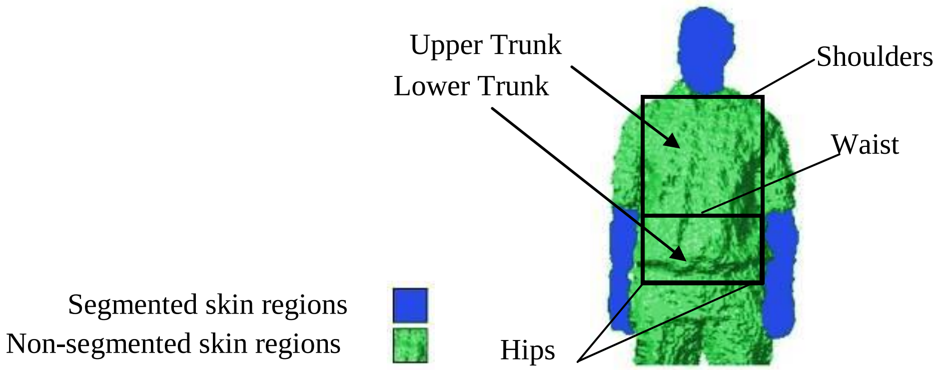


Figure 7 Upper and Lower Trunk Identification Example. [121]

The last body parts to segment are the upper arms. To identify the 3D data regions that represent the upper arms, the previously segmented body parts are removed from the 3D data, leaving only the 3D data corresponding to the lower body below the hips and the upper arms. The lower body region is easily identified as the largest remaining region that has the lowest centroid, hence, this region is also removed from the data, leaving behind the two remaining 3D data regions corresponding to the two upper arms.

3.6 Static Pose Identification

The body part segmentation procedure described above is performed every 0.1 seconds. Bounding boxes are formed around each body part, and the size and centroid of these bounding boxes are tracked to identify a static body pose. Image size normalization is applied to compare normalized centroids and bounding box sizes. If these parameters are within an allowable error of 2.5% when compared to the previous consecutive set of images, the current pose is defined to be the same as those in the previous frames. As previously mentioned, when a pose is held for 4 seconds it is defined as a static pose. Ellipsoids are then fit to the 3D data of each segmented body part of a static pose in order to generate a 3D human upper body model.

3.7 Reverse Tree Structure Ellipsoid Model

Once the 3D data of the body parts have been segmented and a static pose has been identified, ellipsoids are fit to the data. In particular, moment analysis is utilized to estimate the parameters of the ellipsoids similar to [132]. This technique allows for real-time implementation due to its limited computational complexity. By iterating the technique, ellipsoid fitting to the 3D information of the body parts can be achieved more accurately.

The 0th order moment represents the size of 3D points, the 1st order moment identifies the position of the center of the generated ellipsoid, and the 2nd order moment is used to generate orientation information. Equations (7)-(9) formulate the 0th, 1st and 2nd order moments respectively:

$$M_0 = \sum_{p \in P} 1, \tag{7}$$

$$M_i = \frac{1}{M_0} \sum_{p \in P} i_p \; , i \in \{x, y, z\} \text{ and} \tag{8}$$

$$M_{ij} = \frac{1}{M_0} \sum_{p \in P} i_p \, j_p \; , (i, j) \in \{x, y, z\}. \tag{9}$$

Since the 1st order moment represents the center coordinates of the ellipsoid, the following relationship can be defined:

$$(x_P, y_P, z_P) = (M_x, M_y, M_z)\,. \tag{10}$$

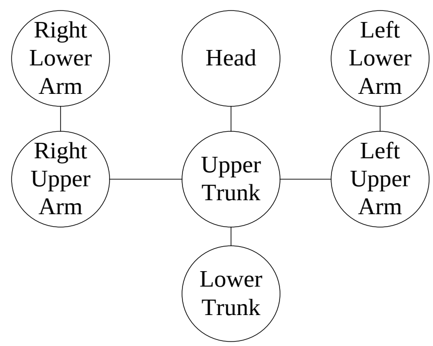
By decoupling the 2nd order moment matrix derived from Equation (9), the lengths and orientation information of the three axes of the ellipsoid can be determined:

$$M = 4\begin{bmatrix} M_{xx} & M_{xy} & M_{xz} \\ M_{yx} & M_{yy} & M_{yz} \\ M_{zx} & M_{zy} & M_{zz} \end{bmatrix} = R_P \begin{bmatrix} \alpha_P^2 & 0 & 0 \\ 0 & \beta_P^2 & 0 \\ 0 & 0 & \gamma_P^2 \end{bmatrix} R_P^T \,. \tag{11}$$

By aligning the elliptical axes with the coordinate axes in Equation (11), solutions for R_{Px}, R_{Py}, R_{Pz} and $\alpha_P, \beta_P, \gamma_P$ can be easily found. The aforementioned technique is iterated until all ellipsoid parameters converge.

The ellipsoids for all the body parts are then connected together to form a full upper body ellipsoid model by applying a reverse tree structure, Figure 8(a). The head and lower arms are taken as the base of the reverse tree structure, to which the other body parts are connected, Figure 8(b). This reverse tree structure ensures that the appropriate orientations of the lower arms are maintained. The ellipsoid model can then be used to determine the degree of accessibility of a person towards the robot using the position accessibility scale of NISA.

(a)



(b)

Figure 8 (a) Ellipsoid model, (b) reverse tree structure of body model.

3.8 NISA Static Body Pose Classification

The static pose information is utilized with the NISA accessibility scale to determine how accessible a person is to a robot. Since body positions are important indicators of naturalistic behaviours, NISA has been effective in demonstrating that the static body pose of one person relative to another person is linked to his/her degree of psychological openness and rapport, and emotional involvement. The position accessibility scale consists of 4 distinct levels, ranging from Level IV (most accessible) to Level I (least accessible).The orientation patterns, toward (T), neutral (N) or away (A), of the upper and lower trunks relative to the robot define the levels of the scale. These accessibility levels are then divided into 3 sub-levels based on the T, N or A arm patterns as previously discussed. Table 4 presents an overview of the static body pose categorization as a function of the trunk and arm patterns. For example, a person whose upper and lower trunks are in a towards orientation (T/T) and who’s arm orientation is away (A) is considered to be in accessibility level IV, 10. Davis performed and analyzed over 50 different interactions to determine the criteria for the body pose coding presented in Table 4. Davis considered replicability and perceptual saliency when determining the coding and found that interrater reliability of the coding using Cohen's Kappa [133] for its nonverbal variables was 0.4-0.85 [119].

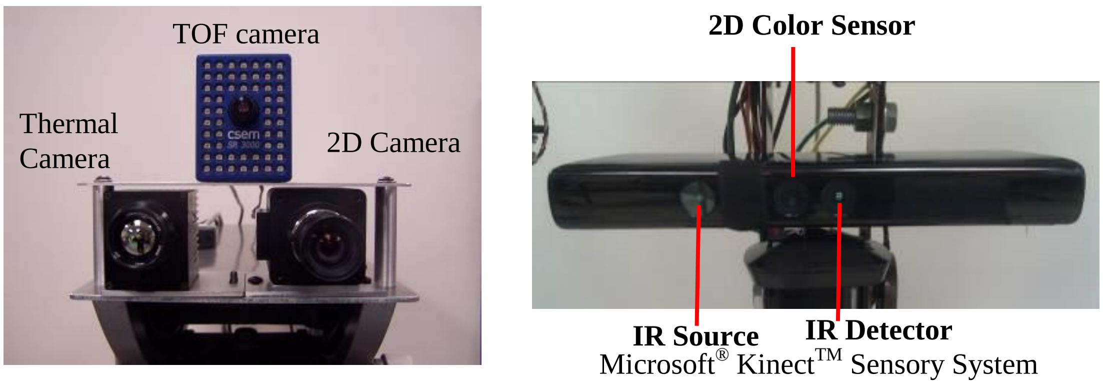
In order to determine the accessibility level of a person towards the robot, the following two step algorithm was developed: (i) Step 1, the ellipsoid parameters are utilized to classify the orientations of the upper and lower trunks, with respect to the reference T, N, and A patterns described above. The criteria in Table 4 is then utilized to classify the accessibility level based on this trunk information; (ii) Step 2, the ellipsoid parameters are used to classify the arm patterns as T, N, and A, to determine the finer-scale accessibility level. This finer position scaling is coded on a 12-point accessibility scale with respect to the trunk orientations, where 1 represents least accessible and 12 is most frontally oriented and toward.

Table 4 Accessibility Levels.

Trunk Orientation	Accessibility Level	Arm Orientation	Finer-Scaling
Upper/Lower trunk: T/T, T/N or N/T	IV	T	12
		N	11
		A	10
Upper/Lower trunk: T/N or N/T except positions that involve upright or forward leans	III	T	9
		N	8
		A	7
Upper/Lower trunk: N/N, A/N, N/A, T/A, A/T	II	T	6
		N	5
		A	4
Upper/Lower trunk: A/A	I	T	3
		N	2
		A	1

3.9 Sensory Systems

Since the application of this accessibility recognition methodology is for a mobile robot engaging in one-on-one socially assistive HRI, it is important that the corresponding real-time sensory systems that are used by the proposed approach be non-contact, compact and placed directly on the robotic platform. Two different sensory systems that can provide both 3D data and skin region information were incorporated with the body language recognition and classification technique in order to investigate the robustness of the technique. The first sensory system consists of a TOF 3D camera to provide 3D data and a thermal camera to provide skin region information, Figure 9(a). A 2D camera is also presented with the sensory system, but it is not directly employed for the proposed automated identification and classification technique, rather it is used to allow an expert coder to verify a person's accessibility level towards the robot during experiments. The second sensory system utilized is the Microsoft® Kinect™ system which consists of a 2D color camera and an IR depth imager, Figure 9(b).



(a) (b)

Figure 9 (a) Thermal camera, TOF Camera and 2D camera, and (b) Kinect™ sensor. [121]

3.9.1 Sensory System 1: Thermal Camera and TOF Camera

The first sensory system uses thermal images for human skin tracking and a TOF camera to generate 3D data of the person interacting with the robot. The thermal camera is a Thermoteknix Miricle 110K thermal camera which achieves detailed real-time thermal images based on an uncooled alpha Silicon 7-14 μm micro-bolometer detector [134]. It can provide thermal images with 384x288 pixel resolution at 50 fps. To generate binary skin data, the pixels in the thermal image that have intensity values above threshold T_h are identified as skin. The TOF camera is a CSEM SwissRanger SR-3000 range camera [135]. It can provide 3D point clouds with a pixel resolution of 176x144 at a frame rate of 25 fps with a range up to 7.5 m. The 2D camera, utilized for system verification, is a Pulnix TMC-6740CL which is a miniature, high-speed color progressive scan CCD camera with a standard 200 fps providing 640x480 pixel resolution [136]. The cameras were calibrated utilizing the Matlab® calibration toolbox [137].

3.9.2 Sensory System 2: Kinect™ Sensor

The Kinect™ sensor consists of a 2D CMOS color camera with a resolution of 640 x 480 pixels and a depth imager with the same resolution. The skin regions are identified in the 2D color information provided by the 2D camera utilizing a YCbCr color space technique [138]. 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 [139]. The operating range of the depth sensor is approximately 50cm-5m. Both the 2D camera and depth imager operate at a maximum frame rate of 60Hz. The Kinect™ sensor was also calibrated utilizing the Matlab® Calibration toolbox.

3.10 Summary

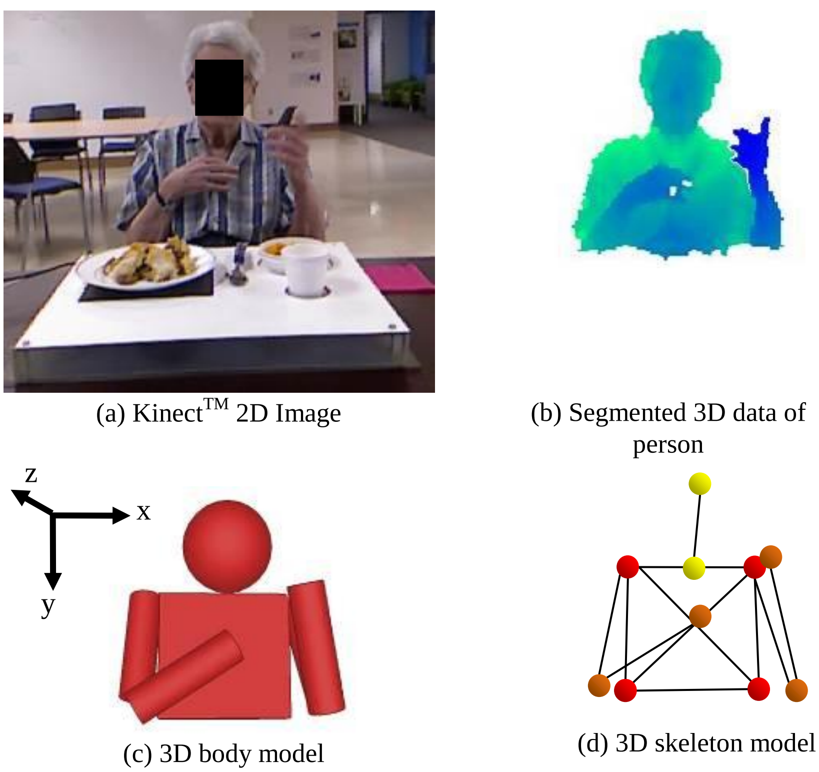
In summary, this chapter has presented a novel affect classification system for use by robots engaged in one-on-one social interactions. The proposed approach is the first technique that recognizes and classifies natural static body language utilizing multimodal sensor data to perform full upper body part segmentation and 3D static body pose identification for automated classification of a person's accessibility level during one-on-one social HRI. The affect classification method is not sensor-dependent and has been implemented with two different sets of sensors, i) a time-of-flight camera and thermal camera, and ii) the 2D and 3D depth imagers of the Microsoft® Kinect™ sensor.

Chapter 4
Detecting Affect from Dynamic Body

In this chapter, an automated dynamic body language recognition and classification system was developed to effectively interpret communicative dynamic nonverbal behaviour of a person during social interactions with a robot. The proposed system has been developed to classify seated body language with respect to valence and arousal. Valence and arousal scales can describe the quality and intensity of both standing and seated affective body language by utilizing a large range of affective states [99, 140]. They are especially effective in describing a person's affective behaviours during dyadic social interactions [141]. In addition, it has been found that valence and arousal better represent experimental and clinical findings than a categorical emotional model (e.g. happy, angry, sad) [142]. The proposed system focuses on seated body language due the large variety of cognitively stimulating and behavioural therapy scenarios for the elderly that are performed while the user is seated, such as a cognitively stimulating card/board games and important ADLs such as meal-eating. Additionally, the majority of residents in long-term care facilities use wheelchairs [143], hence they perform many behaviours while seated.

4.1 Human Body Tracking

The Microsoft® Kinect™ sensor is utilized to obtain depth images of the user. The same technique presented in Chapter 3 is used to extract the 3D data of a person from an interaction scene for dynamic body language tracking. Namely, the 3D data of the person is segmented from the interaction scene utilizing a Mixture of Gaussians (MOG) [122] and a connected component analysis [123] approach. This segmentation technique allows the user to be extracted from sensory data corresponding to other potential moving objects as well as people within the scene. A 3D human upper body model is fit to the segmented data of the user to determine his/her body language. An example Kinect™ 2D image and segmented 3D data of a user engaged in an assistive interaction, in this case a meal-eating activity, with a robot are shown in Figure 10(a) and Figure 10(b), respectively. The corresponding 3D human upper body model is shown in Figure 10(c). During seated assistive scenarios, only the user's upper body language is of interest. Hence, for tracking the upper body in this scenario, the 3D human upper body model consists of a sphere for the head, a rectangular prism for the trunk, and two sets of two cylinders for each arm. A 3D skeleton model, Figure 10(d), is used to connect each of these body parts at the following joints: the neck (joint between the head and trunk), the left and right shoulders (joints between the upper arms and the trunk), and the left and right elbows (joints between the upper and lower arms). The 3D skeleton model also includes five additional points on the boundaries of the 3D body model parts: the head (top of the sphere of the 3D body model), the left and right hips (bottom left and right front corners of the rectangular prism of the 3D body model), and left and right hands (ends of the lower arm cylinders). The i^{th} 3D point of the skeleton model, $p_{i,f} = (p_{x_{i,f}} , p_{y_{i,f}} , p_{z_{i,f}})$, is defined for each frame, *f*, of a body language display. The 2D images from the Kinect™ sensor are used for human baseline coding of the person's affective body language.



(a) Kinect™ 2D Image

(b) Segmented 3D data of person

(c) 3D body model

(d) 3D skeleton model

Figure 10 Example Kinect™ data and models.

Body tracking is performed using the Dynamic Bayesian Network (DBN) technique with a Ray Constrained Iterative Closest Point (RC-ICP) measurement model presented in [144]. The aforementioned 3D body model (including the 3D skeleton) is fit to each frame of 3D user data utilizing the RC-ICP procedure, and the DBN tracks body language through multiple 3D data frames while applying motion and geometric constraints to the updated model configuration. The 3D skeleton model information of a user's body language display is then used to estimate the user's affective state during interactions. The DBN RC-ICP approach is used herein as it provides more accurate joint position estimations when compared to other approaches such as articulated ICP tracking and model-based ray-casting techniques [144].

4.2 Body Language Features

A number of body language features with respect to the overall movement and poses of the body as well as individual body parts are identified using the tracked 3D skeleton model in order to autonomously classify an affective body language display. In particular, body language features for the overall body include vertical and forwards/backwards motions, speed, and expansiveness (the spatial volume used for the body language display). The body language features for individual body parts are defined for the head, arms and trunk, and include vertical and forwards/backwards head positions, opening and closing of the arms, and bowing and stretching of the trunk. Previous studies in psychology have directly linked these body language features to a person's levels of valence and arousal [145-147]. Furthermore, these body language features have also been used to distinctly identify a number of discrete affective states [47]. The descriptions of the features and how they are obtained from the skeleton model are presented in Table 5. Figure 11 shows example body movements and poses for each body language feature.

Table 5 Detected Body Language Features.

Feature	Description	Formulation
Bowing/ Stretching of the Trunk	average trunk lean angle towards or away from the robot during the body language display	$\frac{1}{N}\sum_{f=1}^{N}\left(\arctan\left(\frac{p_{y_{shoulder,f}}-p_{y_{hip,f}}}{p_{z_{shoulder,f}}-p_{z_{hip,f}}}\right)\right),$ (12) where N is the total number of 3D data frames in a body language display, and $p_{shoulder,f}=\frac{1}{2}\left(p_{leftshoulder,f}+p_{rightshoulder,f}\right)$ and $p_{hip,f}=\frac{1}{2}\left(p_{lefthip,f}+p_{righthip,f}\right).$
Opening/ Closing of the Arms	average distance between the hands and the center of the trunk during the body language display	$\frac{1}{N}\sum_{f=1}^{N}\left(\frac{1}{2}\left\Vert p_{lefthand,f}-p_{trunkcenter,f}\right\Vert+\frac{1}{2}\left\Vert p_{righthand,f}-p_{trunkcenter,f}\right\Vert\right),$ (13) where $p_{trunkcenter,f}$ is the centroid with respect to $p_{leftshoulder,f}$, $p_{rightshoulder,f}$, $p_{lefthip,f}$, and $p_{righthip,f}$ points at frame f.
Vertical Head Position	average relative height of the head with respect to the neck during the body language display	$\frac{1}{N}\sum_{f=1}^{N}\left(p_{y_{head,f}}-p_{y_{neck,f}}\right)$ (14)
Forward/ Backwards Head Position	average distance between the head and the neck towards or away from the robot during the body language display	$\frac{1}{N}\sum_{f=1}^{N}\left(p_{z_{head,f}}-p_{z_{neck,f}}\right)$ (15)
Vertical Motion of the Body	average upwards/ downwards movement of the body during the body language display	$\frac{1}{N-1}\sum_{f=1}^{N-1}\left(\frac{1}{S}\sum_{i=1}^{S}\left(p_{y_{i,f+1}}-p_{y_{i,f}}\right)\right),$ (16) where S is the total number of points on the skeleton model.
Forward/ Backwards Motion of the Body	average towards or away movement of the body with respect to the robot during the body language display	$\frac{1}{N-1}\sum_{f=1}^{N-1}\left(\frac{1}{S}\sum_{i=1}^{S}\left(p_{z_{i,f+1}}-p_{z_{i,f}}\right)\right)$ (17)
Expansiveness of the Body	average spatial extension of the body during the body language display	$\frac{1}{N}\sum_{f=1}^{N}\left(\left(\max_i p_{x_{i,f}}-\min_i p_{x_{i,f}}\right)\left(\max_i p_{y_{i,f}}-\min_i p_{y_{i,f}}\right)\left(\max_i p_{z_{i,f}}-\min_i p_{z_{i,f}}\right)\right)$ (18)
Speed of the Body	average velocity of the movement of the body during the body language display	$\frac{1}{N-1}\sum_{f=1}^{N-1}\left(\frac{1}{S}\sum_{i=1}^{S}\left(\frac{\left\Vert p_{i,f+1}-p_{i,f}\right\Vert}{T_{f+1}-T_{f}}\right)\right),$ (19) where T_f is the time at frame f.

(a) Stretching trunk motion

(b) Opening arms motion

(c) Head moving upwards

(d) Head moving forwards

(e) Overall upwards body motion

(f) Overall forwards body motion

(g) Increasing expansiveness body motion

(h) High speed body motion

Figure 11 Example body language displays for each feature.

4.3 Affect Estimation from Body Language

The aforementioned body language features are utilized as inputs into an automated affect estimation system in order to classify valence and arousal levels of a person during socially assistive HRI. Valence represents the level of pleasantness (or unpleasantness) that is felt with respect to a stimulus (the robot and interaction items), while arousal refers to the level of intensity of affect [148]. This chapter investigates different learning algorithms for affect estimation in order to be able to effectively identify dynamic affective body language displays during HRI. Namely, the following classes of learning techniques are implemented: 1) probabilistic techniques, 2) linear techniques utilizing probabilistic information, 3) decision trees, 4) lazy learning algorithms, 5) meta-classifiers, 6) neural networks, and 7) non-linear models. Table 6 presents a summary of the seven techniques implemented within the aforementioned classes for estimating both valence and arousal.

Naïve Bayes was chosen as the probabilistic technique due to its robustness to irrelevant features [149], allowing body language features that do not contribute to the classification of certain affective states to be omitted. Logistic Regression is chosen as the linear technique due to its ability to handle various data types that do not need to be normally distributed, linearly related or of equal variance [150]. This allows Logistic Regression to effectively classify body language features vectors that contain different types of data, namely features that will result only in positive values (e.g. speed of the body), features that could be positive or negative(e.g., vertical motion of the body) and features that may have large differences in magnitude (e.g., expansiveness of the body and vertical head position). Random Forest was chosen as the decision tree technique due to its robustness to over-fitting and outliers [151], hence, being able to deal with outlier body language displays with feature values that do not match the majority of other displays for the same affective state. Robustness to over-fitting makes Random Forest well suited to generating classifications based on a variety of different individuals. The K-nearest neighbour (KNN) technique was chosen as the lazy learning technique due to it retaining all training data for classification and its ability to estimate a complex target function [152]. This allows KNN to correctly classify affective body language displays even when there are complex non-linear relationships between body language features and affective states. Adaboost was chosen as the meta-classifier due to its ability to re-weight the base classifier’s misclassified samples and generate an updated classifier [153]. Adaboost is able to accurately classify a wider range of training samples than the base classifier for Naïve Bayes, accounting for variations in body language displays across different individuals when classifying the same appropriate affective state. A Radial Basis Function (RBF) network was chosen as the neural network technique and support vector machines (SVMs) were chosen as the non-linear technique due to their ability to respond well to feature vectors that have not been utilized in training [152,153]. Both can effectively classify body language displayed by individuals not used in the training data. Additionally, SVMs are effective at dealing with smaller sets of data for some classes [154]. SVMs can generate effective class boundaries even if a smaller number of training body language displays are collected for certain affective states, which may occur when collecting natural affective body language displays during socially assistive interactions.

4.4 Summary

In summary, this chapter presents a novel automated affective body language recognition and classification system that will allow socially assistive robots to interpret natural seated displays of dynamic affective human body language during socially assistive HRI. The proposed system utilizes 3D data from an onboard Kinect™ sensor to recognize and track body language features using the 3D skeleton of a person’s upper body. The features can then be classified into valence and arousal values using a variety of learning-based classifiers.

Table 6 Summary of Investigated Learning Techniques.

Technique Class	**Technique**	**Description**
Probabilistic	Naïve Bayes	A feature vector is assigned to a classification category that has the highest probability for that feature vector [155]. The probability that a feature vector belongs to a certain category is determined by the product of the probabilities of each feature belonging to that category.
Linear using probabilistic information	Logistic Regression	A feature vector is assigned to a classification category with the highest probability for that feature vector using a logistic function. The logistic function determines the probability of a category based on a weighted linear combination of the features in a feature vector [156]. The logistic function parameters and linear weights are identified utilizing training data.
Decision Trees	Random Forest	A large number of decision trees are generated, each based on a random subset of training data, to determine the classification category of a feature vector [157]. The feature vector is assigned to the category that is the most frequent classification category obtained from all the trees.
Lazy Learning	*K*-Nearest Neighbor (KNN)	The distance from a feature vector to every training feature vector is determined. The classification category that represents the majority of the *k* closest training feature vectors is assigned to the feature vector of interest [158].
Meta-classifiers	Adaptive Boosting (Adaboost) with Naïve Bayes	A feature vector is assigned a classification category corresponding to the weighted linear combination of a number of Naïve Bayes classifiers [159]. The weight of each classifier is determined based on the error rate obtained from classifying the training data.
Neural Networks	Radial Basis Function (RBF) Network	A two layer (a hidden node layer and a weighted linear combination layer) feed forward neural network is used to classify a feature vector [160]. At each hidden node in the network an RBF is calculated for each input feature vector. The output classification node with the largest weighted sum of RBF results is defined to be the assigned category of the input feature vector. The RBF parameters and linear weights are identified utilizing training data.
Non-Linear Models	Support Vector Machines (SVM)	One or more hyperplanes (dividing surfaces between classification categories) are determined utilizing non-linear kernel functions and training data [161]. A feature vector is assigned a classification category by determining which category region (bounded by hyperplanes) the feature vector is located within.

Chapter 5
Designing Affective Body Language for a Socially Assistive Human-Like Robot

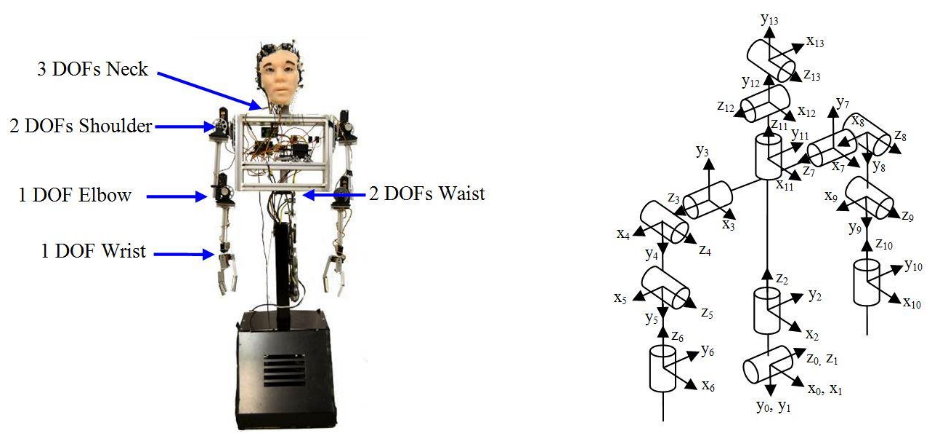
This chapter aims to identify the appropriate emotional body language for a human-like robot, Brian 2.1, to display during natural one-on-one social interactions with a person. Namely body language displays are designed to allow Brian 2.1 to communicate a variety of affective states. Uniquely in this thesis, the movements and postures of the robot are adapted from body language displays, identified in behavioural and psychology research, to correspond to specific human affect displays. This work implements the body movements and postures defined by Wallbott [47] and de Meijer [48] for a number of different affective body language displays.

5.1 Social Robot Brian 2.1

The human-like robot Brian 2.1 has similar functionalities to a human from the waist up, Figure 12(a). The dimensions of the upper body of the robot have been modeled after a male volunteer. The robot is able to display non-verbal body language via: a) a 3 DOFs neck capable of expressing realistic head motions such as nodding up and down, shaking from side to side and cocking from shoulder to shoulder, (b) an upper torso consisting of a 2 DOFs waist allowing it to lean forward and backwards as well as turn side to side, and (c) two arms with 4 DOFs each: 2 DOFs at the shoulder, 1 DOF at the elbow and 1 DOF at the wrist. Utilizing these body parts, the robot is capable of displaying various human-like body movements and postures.

5.2 Affective Body Language Features

As previously mentioned, since both body movements and postures are important cues for recognizing emotional states displayed by an individual, this chapter focuses on defining emotional body language for the robot Brian 2.1 that encompasses both these characteristics. This body language should be consistent with emotions that a robot would display during social HRI scenarios. The body language classification of Wallbott [47] and de Meijer [48] are utilized to generate body language corresponding to the emotions of sadness, elated joy, anger, interest, fear, surprise, boredom and happiness. This set of eight emotions was chosen as it provides a large variation across both the valence (positive and negative feelings) and arousal (level of activity) dimensions of affect. For example, sadness represents negative valence whereas elated joy represents positive valence; and boredom represents low arousal whereas surprise represents high arousal. Furthermore, these emotions are included within a group of emotions that psychologists define as social emotions, [162-164]. Namely, social emotions which can also include the basic emotions of happiness, sadness, fear and anger serve a social and interpersonal function, where an individual's relationship to another individual can be the central concern for these emotions [165-167]. Hence, these emotions involve the presence of a (real or virtual) social object which may include another person or a social constructed self [168]. The set of eight emotions that was chosen, herein, can be used by the robot to engage in social communication with a person in order to accomplish different interaction goals such as, for example, obtaining compliance or gathering information.



(a) Brian 2.1 (b) Kinematic model of Brian 2.1

Figure 12 13 DOFs Human-like Social Robot Brian 2.1. [169]

The body language descriptors used for the different emotions are presented in Table 7. The emotions of sadness, elated joy, boredom and happiness are derived from body movements defined by Wallbott [47] and the other four emotions of anger, interest, fear and surprise are derived from de Meijer's work [48]. Body language classification was taken from both these works in order to accommodate the range of proposed emotions for Brian 2.1. The emotional body movements and postures chosen can be achieved based on the robot's mobility specifications and include upper trunk, head and arm movements as well as the overall movement quality. Trunk movement is classified as either the stretching or bowing of the upper trunk, which the robot emulates by leaning forwards or backwards at the waist. The movement of the head consists of facing forwards, tilting backwards or facing downwards and is achieved via the robot's 3 DOFs neck. The arm motions are defined as: i) hanging- when resting at the sides of the robot, and ii) opening/closing- for opening, the arms start near the center of the robot and move outwards away from the body, while closing consists of the opposite motion. The overall direction of movement is also described as forwards, backwards, upwards and downwards based on the motion of the trunk, arms and head of the robot in these directions. The movement quality represents the overall speed, size and force of movements and is divided into three main categories [47]: 1) movement dynamics- which refers to the energy, force or power in a movement; 2) movement activity- which refers to the amount of movements, and 3) expansive or inexpansive movements- which refer to the large or small spatial extension of the robot's body.

Table 7 Body Language Descriptors for Different Emotions.

Emotion	Body Movements and Postures
Sadness	Bowing trunk, head forward, hanging arms, and low movement dynamics.
Elated joy	Stretching trunk, opening arms, overall upward motions, and high movement activity and dynamics with expansive movements.
Anger	Bowing trunk, high movement activity, and high movement dynamics.
Interest	Stretching trunk, opening arms, overall upward and forward motions, and low movement dynamics.
Fear	Bowing trunk, closing arms, overall backward motion, and high movement dynamics.
Surprise	Stretching trunk, overall backward motion, and high movement dynamics.
Boredom	Bowing trunk, head tilted back, hanging arms, and low movement activity with inexpansive movements.
Happiness	Stretching trunk, head forward, arms hanging, and low movement dynamics.

In order to implement the emotional body language descriptors in Table 7, the kinematic model for Brian 2.1 (shown in Figure 12(b)) is utilized. For example, to implement the descriptors for elated joy, the following joints are used. The revolute joint 1 rotates the robot's trunk to an upright position, where the trunk is perpendicular to the ground to represent a stretched trunk posture. The two shoulder joints (joints 3 and 4, and 7 and 8) and elbow joints (joints 5 and 9) of each arm are used to move the arms of the robot in an upwards and outwards direction to mimic opening of the arms. Joint 12 is used to tilt the head back. The combination of the trunk, arms and head motion represents the overall upward motion of the robot. High movement activity is achieved by repeating the upwards motions several times. High movement dynamics are achieved by high joint velocities. Expansive movements increase the spatial workspace of the robot during the display of body language and are implemented through the motion of opening the arms as well as the rotating of both the trunk and head from left to right using joints 2 and 11. Video frames showing the robot's body language displays, with descriptions, are shown in Figure 13.

5.3 Summary

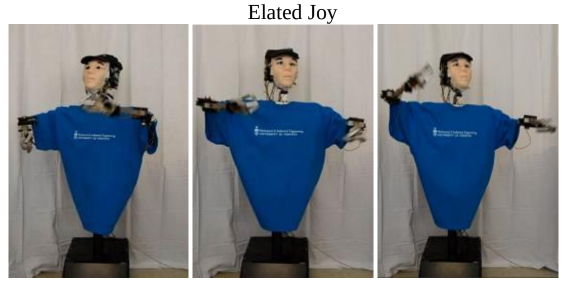
In summary, this chapter presents robotic affective communication as displayed through body language designed for social HRI scenarios. Uniquely, the emotional body language for the human-like social robot Brian 2.1 were based on body movement and posture descriptors identified in human psychology and behavioural research. Additionally, distinct from other robot body language studies in the literature, this work focused on the use of social emotions that could be the causation of interpersonal factors during social HRI. The body language descriptors utilized for the robot are based on trunk, head and arm movements as well as overall movement quality.

Sadness

Frame 15 Frame 47 Frame 96

The trunk is rotated forward to represent a bowing trunk posture, while small arm movements are implemented. The head is slowly rotated from side to side as well as down. Low joint velocities are used to represent low movement dynamics.

Elated Joy



Frame 9 Frame 37 Frame 45

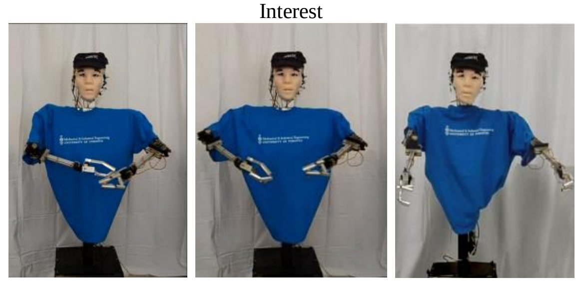
The trunk is in an upright position to represent a stretched trunk posture. Upwards and outwards motions are generated for the arms, while the trunk and head are rotated from side to side for expansive movements. The head is also tilted up. High joint velocities and frequency of motions represent high movemement dynamics and activity.

Anger

Frame 35 Frame 55 Frame 90

The trunk is rotated forward to represent a bowing trunk posture, while high joint velocities are used for frequent downward motions of the arms and head to mimic high movement dynamics and activity.

Interest



Frame 5 Frame 45 Frame 67

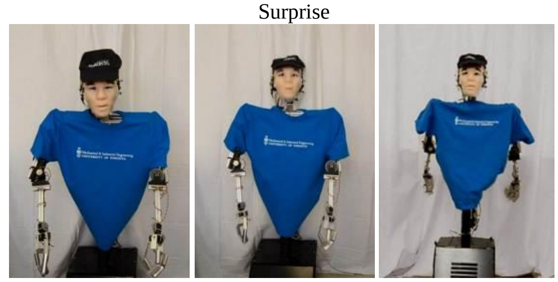
The trunk is in an upright position to represent a stretched trunk posture. Forward motions are represented by the opening of the arms away from the robot. Low joint velocities create low movement dynamics.

Fear

Frame 38 Frame 52 Frame 71

The trunk is rotated forward to represent a bowing trunk posture while also being rotated to the robot's left and away from the camera to represent a backwards motion. The head and arms are brought in closer to the trunk. High joint velocities generate high movement dynamics.

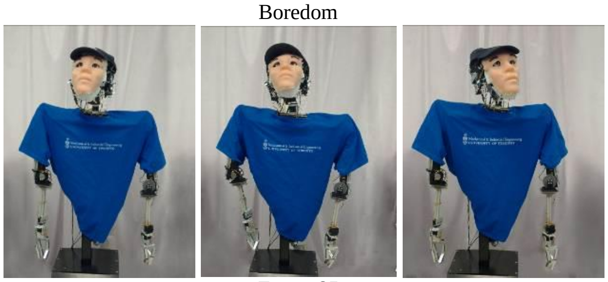
Surprise



Frame 63 Frame 93 Frame 123

The trunk is rotated from a forward position to an upright position to represent stretching of the trunk. The arms are rotated behind the trunk to represent a backwards motion. High joint velocities are used to represent high movement dynamics.

Boredom



Frame 21 Frame 85 Frame 145

The trunk is rotated forward to represent a bowing trunk posture. The head is tilted up and rotated slowly from side to side. The arms are hanging and move slightly at the sides of the robot to show inexpansive movements. Low frequency motions create low movement activity.

Happiness

Frame 24 Frame 78 Frame 158

The trunk in an upright position to represent a stretched trunk posture. The head is forward and the arms are moving slowly at the sides of the robot to represent low movement dynamics.

Figure 13 Example Frames of Emotional Body Language Displayed by Brian 2.1 for the Eight Emotions. [169]

Chapter 6
Socially Assistive Human-Robot Interaction Experiments

As one of the main applications of this work is social HRI, this chapter uniquely investigates and compares how elderly users interact with and perceive Brian 2.1 in two different socially assistive scenarios in long-term care: i) playing the memory card game in a public setting, and ii) one-on-one meal-eating with the robot in a private room. These two scenarios are investigated to determine if Brian 2.1 can be a potentially effective resource at engaging and assisting the elderly with different types of ADLs in various settings in long-term care facilities.

6.1 Design of Socially Assistive Robot Behaviours

In order to promote natural interactions between elderly users and Brian 2.1, the robot has been developed to incorporate human-like communication modalities, i.e., body language, facial expressions, speech and vocal intonation. As discussed in Chapter 5, Brian 2.1 can display body language utilizing its: 1) 2 degrees of freedom (DOF) waist that allows it to lean forward and backwards as well as turn left and right; 2) two 4 DOF arms that can mimic human gestures such as pointing and waving; and 3) a 3 DOF neck that allows Brian 2.1's head to turn left and right, tilt forward and backwards, and tilt from side to side. Additionally, a 4 DOF face permits Brian 2.1 to display expressions such as smiling for happiness and frowning for sadness as well as a neutral facial expression. An onboard speaker is utilized to play the robot's synthesized male voice which uses a combination of speech and vocal intonation. Vocal intonation allows Brian 2.1 to produce the corresponding neutral, sad and happy voices by varying the pitch and speed. Together these communication modalities are utilized to display Brian 2.1's assistive emotional behaviours.

The behaviours of Brian 2.1 for each activity are divided into 5 groups: 1) instructions: prompts to inform a user of the next step of an activity (displayed with neutral or sad emotions); 2) encouragement: prompts with positive reasoning tactics to promote engagement in and completion of the activity (displayed with neutral or happy emotions); 3) celebration: congratulating a user on completing a task in the activity or the overall activity itself (displayed with happy emotions); 4) help: a descriptive set of directions for users if they get stuck at a particular task during the activity (displayed with neutral, happy or sad emotions); and 5) general positive statements – social utterances to enhance the overall activity experience which include jokes, user and scenario compliments, and greetings (displayed with happy emotions). The choice of emotional display for the robot is directly related to the level of engagement of the user. Initial instruction and help behaviours are displayed with a neutral emotional state. If the user has been engaged in the activity for a significant amount of time or has completed activity tasks, Brian 2.1 displays a happy emotional state. The sad emotional state is only used by Brian 2.1 to reengage users who have become distracted from the activity.

6.1.1 Memory Card Game

The memory card game is played by users flipping over and matching each pair of picture cards. The game starts with 8 pairs of picture cards randomly placed face down on a table in a 4 by 4 grid (Figure 14), then a user flips over two cards at a time until all picture cards are matched. During the game, Brian 2.1 identifies and monitors the location of flipped over cards via an overhead camera by tracking unique SIFT features [170] on each pair of picture cards. The memory card game allows the robot to adapt its level of assistance during the game based on the cognitive abilities of a user by changing its behaviours. This, in turn, allows individuals with different cognitive abilities to play the game with Brian 2.1.

Figure 14 Example setup of picture cards.

Examples of Brian 2.1's assistive behaviours for the cognitively stimulating memory card game with respect to the five aforementioned groups include: 1) instructions – Brian 2.1 prompting the user to flip over cards by saying "Please flip over two cards." while displaying a neutral facial expression and pointing to the card set-up; 2) encouragement – Brian 2.1 encouraging the user to continue the game when matching pairs have not been found "Those are fascinating cards that you have flipped over, but, unfortunately they do not match. Please turn the cards back over and try again! You are doing great, let's keep going!" while displaying a happy facial expression; 3) celebration – Brian 2.1 congratulating the user on a successful pair of matching picture cards "Amazing job, those cards are a perfect match! You can remove them from the game." while displaying a happy facial expression and raising its arm up in a celebration gesture; 4) help when user is distracted - Brian 2.1 providing a hint to the user "I think you may want to flip over this card." while displaying a sad facial expression and pointing to a card location; and 5) general positive statement – Brian 2.1 complimenting the interactions with the user "I am having a lot of fun playing the memory card game with you." while displaying a happy facial expression. Figure 15 shows example instruction and celebration robot behaviours during the memory card game.

(a) Brian 2.1 instructing user to turn over a card

(b) Brian 2.1 celebrating a match with the user

Figure 15 Example robot behaviours during the memory card game.

6.1.2 Meal-Eating Activity

During the one-on-one meal-eating activity, Brian 2.1 assists users to eat their food by providing appropriate prompts and encouragements. The user's meal-eating actions are monitored by Brian 2.1 by utilizing a meal-tray with embedded weight sensors as well as a utensil tracking system. Figure 16 shows the meal-tray setup for the meal-eating activity. The meal-tray estimates the user's consumption of food or drink for each tableware item utilizing the embedded weight sensors. The 3D location of the utensil is determined during meal-eating by monitoring the infrared (IR) LEDs mounted at the top of its handle via stereo IR cameras placed on the robot's shoulder. Tracking the utensil allows the robot to determine if a user is using the utensil to take food from the meal-tray and place it in their mouth. Brian 2.1 uses the meal-tray and utensil tracking sensory data to assist the user in eating a meal based on a meal-plan prepared by a caregiver.

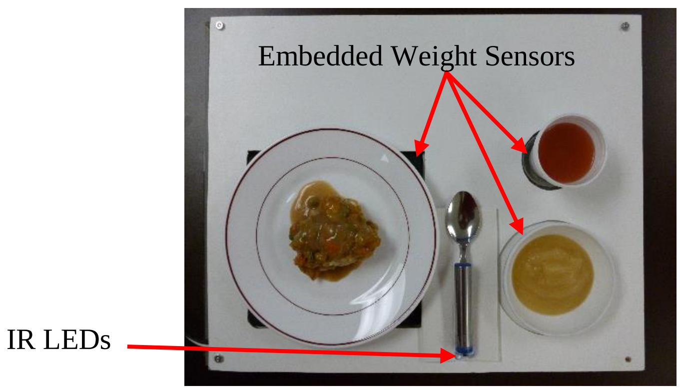
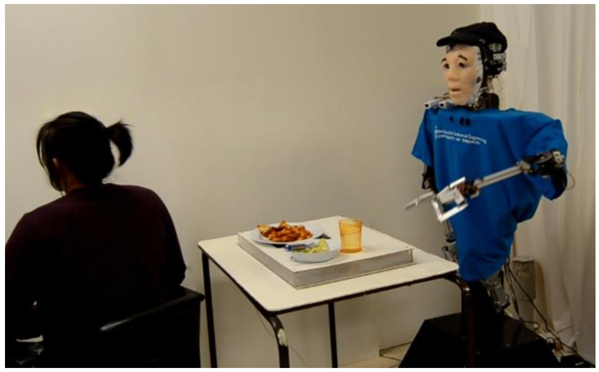


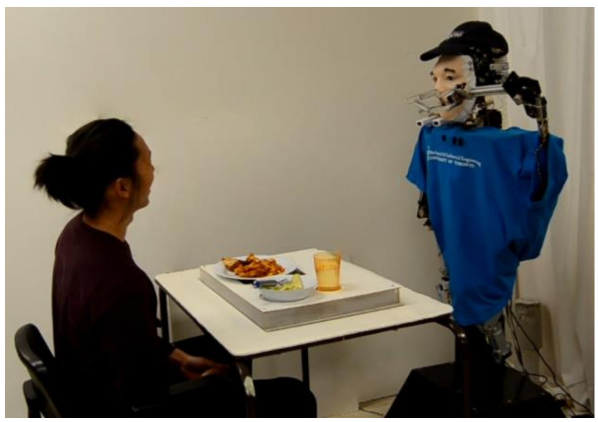
Figure 16 Example meal-tray setup.

Examples of Brian 2.1's assistive behaviours for the meal-eating activity include: 1) instructions – Brian 2.1 prompting the user to pick food up the utensil "Please pick up the spoon and take some food from the main dish." while displaying a neutral facial expression and pointing to the utensil; 2) encouragement – Brian 2.1 encouraging the user to continue eating the meal "The main dish looks delicious. You should have some more!" while displaying a happy facial expression; 3) celebration – Brian 2.1 congratulating the user on completing the meal "Great job, you finished your meal!" while displaying a happy facial expression; 4) help when user is distracted – Brian 2.1 reorienting a disengaged user back to the tray "Please have some juice that is here on your tray." While displaying a sad facial expression and pointing to the beverage cup; and 5) general positive statement – Brian 2.1 telling a joke "What do you give to a sick lemon?

Lemon-aid!" while displaying a happy facial expression and then laughing into its hand. Figure 17 shows Brian 2.1 displaying example help and general positive statement behaviours during the meal-eating activity.



(a) Brian 2.1 providing a help behaviour (while sad) when a user is disengaged



(b) Brian 2.1 covering its mouth while laughing after telling a joke

Figure 17 Example robot behaviours during the meal-eating activity.

6.2 Experiments

Two user studies were conducted to investigate how elderly users at a long-term care facility interact with Brian 2.1 during different assistive scenarios. For both studies Brian 2.1 was controlled by a human operator in a wizard of oz fashion, namely the operator was located in room away from the interaction scene, out of sight of any participants. The first user study took place in an open public space on the 1st floor of the facility, where it would be possible for a number of passersby to stop and play the memory card game with the robot as well as interact amongst themselves. The second user study took place in a designated room where users were invited to eat a meal one-on-one with Brian 2.1.

For both study settings, the following parameters were investigated: 1) length of time that a user interacted with Brian 2.1; 2) the presence of engagement indicators including time spent looking towards the activity or Brian 2.1, manipulating activity items, and/or the presence of utterances towards the robot; 3) compliance with respect to the robot's suggested behaviours; and 4) acceptance and attitudes towards Brian 2.1.

With respect to acceptance and attitudes towards Brian 2.1, a post-study questionnaire was administered after each user study. The questionnaire was adapted from the Almere technology acceptance model for socially interacting agents [171]. The adapted questionnaire consists of 9 different constructs, each consisting of one or more statements, as shown in Table 8. Each participant was asked to rate his/her agreement with each statement on a 5-point Likert scale, with 5 representing strongly agree with the statement and 1 representing strongly disagree. Participants were also asked to identify which robot characteristics they liked. In addition, the questionnaire also collected demographic information and the participants' previous technology experience with computers. Participants were also welcome to provide any additional feedback they may have had on Brian 2.1.

Table 8 Statements and Constructs of the Adapted Almere Model.

Statement	Construct
1. I enjoy the robot talking to me	Perceived Enjoyment (PENJ)
2. When interacting with the robot I felt like I'm talking to a real person 3. It sometimes felt as if the robot was really looking at me 4. I can imagine the robot to be a living creature 5. Sometimes the robot seems to have real feelings	Social Presence (SP)
6. I would trust the robot if it gave me advice 7. I would follow the advice the robot gives me	Trust (TR)
8. I think it's a good idea to use the robot 9. The robot would make my life more interesting	Attitude Towards Using the Robot (ATT)
10. I consider the robot a pleasant conversational partner 11. I feel the robot understands me 12. I think the robot is nice	Perceived Sociability (PS)
13. I find the robot easy to use	Perceived Ease of Use (PEOU)
14. I think I'll use the robot again	Intent to Use (ITU)
15. I think the robot is useful to me 16. I think the robot can help me with many things	Perceived Usefulness (PU)
17. I think the robot can help me with what I need	Perceived Adaptability (PAD)

6.2.1 Memory Game User Study

6.2.1.1 Methods and Participants

The public setting user study took place in a large atrium at the local long-term care facility. The goal was to investigate the interaction of elderly individuals of varying cognitive ability with the socially assistive robot Brian 2.1 during the memory card game.

Brian 2.1 was placed in the atrium for a two-day duration. During this time, Brian 2.1 would introduce itself to passersby and ask them to join the robot in playing the memory card game. Researchers were also present to monitor the interactions between Brian 2.1 and participants, administer the questionnaire, and answer questions regarding the robot.

Forty participants with varying cognitive abilities including mild cognitive impairment, mild Alzheimer's disease, and normal cognitive control interacted with Brian 2.1 through the memory game. The participants' ages ranged from 57 to 100 years old. Figure 18 shows an example interaction between the elderly and Brian 2.1.

Figure 18 Example memory card game interaction during the public setting user study.

6.2.1.2 Results

The participants, on average, interacted in the memory game with Brian 2.1 for 12.6 minutes. In general, participants played at least one full memory card game with the robot. Participant engagement was categorized into the amount of time a participant was engaged with Brian 2.1 or in the memory card game during the interaction: 1) all of the time; 2) some of the time; and 3) none of the time. A similar categorization was also used for participant compliance, the categories, herein, were defined based on the number of times that a participant complied with Brian 2.1's instructions, encouragement and help behaviours: all of the time, some of the time and none of the time. The engagement and compliance results are shown in Table 9. The one participant who did not comply with any of the robot's prompts had interference with his hearing aid whenever Brian 2.1 spoke. This may have been due to the use of an amplifier to increase the volume of Brian 2.1's voice for the large public place. Six participants spoke directly to the robot, asking it about the card game and how well they were playing. Thirty-three participants smiled in response to the robot's display of a happy facial expression or laughed directly at the robot's jokes. In addition to having a large number of participants interacting with the robot, it was observed that having the robot in such a public setting encouraged social interactions amongst the older adults themselves as well as with caregivers.

Table 9 Engagement and Compliance Results.

Engagement and Compliance Categorization	**Number of Participants that were Engaged**	**Number of Participants that Complied**
All of the time	33	35
Some of the time	7	4
None of the time	0	1

Twenty-two participants (14 female and 8 male) completed the post-study questionnaire. The descriptive statistics for the adapted Almere model are presented in Table 10. Cronbach's alpha (Table 10) was utilized to determine the inter-reliability between statements for each construct that has more than one statement. Constructs with Cronbach's alpha values of 0.5 or greater [172] were further analyzed. All of the constructs with more than one statement obtained alpha values greater than or equal to 0.5. The mean participant ratings for all the constructs were all greater than a neutral Likert scale rating of 3. These results indicate that, on average, the participants enjoyed interacting with and had positive attitudes towards Brian 2.1, felt the robot had a social presence, perceived it to be social and adaptable to their needs, and trusted the robot. They also perceived Brian 2.1 to be easy to use as well as useful, and had intent to use the robot again. The robot characteristics were ranked based on the number of participants that stated they had liked each characteristic. The results are presented in Table 11. As shown in the table, the most-liked characteristic for Brian 2.1 is its ability to express emotions utilizing both facial expressions and vocal intonation.

Table 10 Descriptive Statistics for the Adapted Almere Model.

Construct	Minimum	Maximum	Mean	Standard Deviation
PENJ	4.0	5.0	4.65	0.49
SP (alpha = 0.62)	1.0	5.0	3.46	1.39
TR (alpha = 0.86)	1.0	5.0	3.53	1.32
ATT (alpha = 0.64)	1.0	5.0	4.53	0.89
PS (alpha = 0.5)	1.0	5.0	4.37	0.96
PEOU	2.0	5.0	4.53	0.79
ITU	2.0	5.0	4.53	0.94
PU (alpha = 0.84)	1.0	5.0	3.44	1.50
PAD	1	5	3.59	1.41

Table 11 Most-Liked Robot Characteristics.

Robot Characteristic	Number of Participants That Liked Robot Characteristic
1. Expressing emotions through vocal intonation and facial expressions	18
2. Human-like voice	17
3. Life-like appearance and demeanor	17
4. Providing companionship	15

Table 12 summarizes the participants' prior experiences with computers. The participants had varying computer skills ranging from no experience to advanced, with the majority of them (i.e., 14 participants) being at least at the beginner's level.

Table 12 Prior Experience with Computers.

Experience Level	Number of Participants
No experience	8
Beginner (email, use simple programs)	2
Intermediate (internet, chat)	1
Advanced (editing documents, using complex programs)	11

Spearman's ρ was used to determine if a correlation exists between computer experience and the participants' ratings of PEOU. A ρ of 0.237 was identified, which showed no significant correlation between these two factors for the participants for an α = 0.05. Namely, it was found that all the participants, regardless of their computer experience, found the robot easy to use.

6.2.2 One-on-One Meal-eating Activity

6.2.2.1 Methods and Participants

The one-on-one meal-eating user study took place in a private room at the local long-term care facility. Similar to the public setting study of the memory card game, the goal of this user study was to investigate how elderly individuals interact with Brian 2.1, however, in this case during the meal-eating activity rather than in a memory game.

Participants were invited by staff at the long-term care facility to eat two lunch-time meals with the robot. Each participant was introduced to Brian 2.1 and its capabilities were discussed with the participant before the participant interacted one-on-one with the robot for the first meal. Eight healthy elderly residents (5 female and 3 male) joined Brian 2.1 for lunch on two separate days during the course of a week. The ages of the participants ranged from 82 to 93 years old. Figure 19 shows an example interaction during the meal-eating activity.

Due to the controlled environment of the one-on-one experiments, videos of the interactions were recorded to investigate user behaviours in more detail than the experiments performed in an open public space. In particular, engagement in the interaction for the meal-eating activity was defined by visual focus of attention towards the robot or meal, manipulation of the utensil and cup, and verbal dialogue towards the robot. Compliance is defined by the participants' cooperative actions with respect to Brian 2.1's prompting behaviours that were performed within 2 minutes of the robot's prompt.

Figure 19 Example meal-eating interactions.

6.2.2.2 Results

The results for each of the engagement parameters and the total engagement time for each participant are provided in Table 13. During the two meals, it was found that participants had visual focus of attention either toward the robot or meal for an average of 98% of the total interaction time. The participants spent the remaining 2% of interaction time looking around their environment. Sixty-nine percent of the total interaction time was spent manipulating meal items. On average it was found that the participants spoke 36 utterances to the robot during the two meals, even though they were initially told that the robot was not able to understand verbal communication. From Table 13, it can be seen that longer interaction times also resulted in more social interactions with the robot. In Table 14, the distribution of the utterances of the participants with respect to the robot's behaviours is presented. The majority of utterances were stated after the robot provided positive statements followed by the robot encouraging the participants to eat. It can also be seen that participants stated utterances to Brian 2.1 even when the robot was not displaying any behaviour.

Table 13 Engagement Indicators.

Participant	Total Interaction Time (Minutes)	Visual Focus of Attention Toward the Robot or Meal (Percentage of Total Interaction Time)	Time Spent Manipulating Meal Items (Percentage of Total Interaction Time)	Total Number of Utterances Toward Robot
P1	55.20	100%	71%	86
P2	9.93	95%	84%	20
P3	12.35	97%	85%	24
P4	13.47	99%	42%	23
P5	13.77	99%	64%	34
P6	22.00	98%	78%	30
P7	10.88	98%	81%	21
P8	17.1	96%	43%	51
Average	**19.34**	**98%**	**69%**	**36**

Table 14 Distribution of Participant Utterances With Respect to Robot Behaviour.

Robot Behaviour Type	Example Robot Behaviour	Number of Participant Utterances Toward Robot	Example Participant Utterances Toward Robot
Greeting	"Hello again! Today's menu includes pasta, apple sauce, and juice. Please have lunch with me." (waves while in a happy emotional state)	25	"Thank you. How are you? I have missed you since last week."
Encourage to obtain main dish	"The main dish looks delicious. You should pick up some food with your spoon."(while in a happy emotional state)	26	"I will. Thank you very much."
Encourage to obtain side dish	"The side dish looks very tasty. Why don't you try some?" (while in a happy emotional state)	12	"It's too bad you can't have some, it's pretty good Brian."
Encourage to obtain drink	"You should try some of your beverage. It looks refreshing." (while in a happy emotional state)	32	"Yes I am going to have some."
Encourage to eat	"What you have on your spoon looks like it will taste really good. Please take a bite." (while in a happy emotional state)	33	"I must admit it is different but very tasty."
Encourage to drink	"The drink in your hand looks delightful. Why don't you take a sip?" (while in a happy emotional state)	5	"You know what they gave me Brian? I think that they gave me some cranberry juice."
Positive statements	"I really like your company; I hope we can do this more often." (while in a happy emotional state)	82	"I hope so. I hope I see you again. And see I never forgot your name and I was looking forward to meeting you again."
Joke	"Why did the cookie go to the doctor?" "She was feeling crummy!" (robot laughs and puts one hand in front of its mouth)	18	(chuckles) "Very funny Brian."
Valediction	"Excellent, you have finished your meal. Thanks for spending your lunch with me. (while waving goodbye in a happy emotional state)	28	"I hope to see you again in the future."
No-robot behaviour	Not Applicable	28	Participant commented to Brian on his late arrival to the experiment: "Sorry I was late Brian." Participant asked Brian about interactions with other people: "Have you seen anyone else today?"

Table 15 shows the total number of prompts provided by Brian 2.1 during the interactions with each participant and participant compliance with these prompts. Only one participant had a compliance rate of more than one standard deviation (σ=9.1%) from the overall mean (μ=87%); P2 at 67%. Therefore, the remaining seven participants complied with, on average, 87% of the robot's prompting behaviours. The low rating of compliance (67%) for P2 is due to the fact that this participant did not like the taste of the food in her main dish during her first meal with the robot, which became apparent when she told this to a research team member after she finished her interaction. When the robot prompted her to eat her main dish, she would instead eat the side dish as she liked the taste of that dish.

Table 15 Compliance Indicators.

Participant	Total Number of Prompts by Robot	Prompts Followed by Participant (Percentage of Total Number of Prompts)
P1	28	96%
P2	9	67%
P3	11	91%
P4	13	85%
P5	12	83%
P6	18	94%
P7	10	90%
P8	18	89%
Average	**15**	**87%**

The descriptive statistics for the adapted Almere model are presented in Table 16. The Cronbach's alpha for each construct are also shown in Table 16. All of the constructs with multiple statements except for PU obtained a Cronbach's alpha greater than or equal to 0.5, and therefore PU was not further analyzed. All but the SP and PU constructs had mean participant ratings greater than the neutral Likert scale rating of 3. These results indicate that on average the participants enjoyed interacting with and had positive attitudes towards Brian 2.1, perceived the robot to be social and adaptable to their needs, and trusted the robot. They also perceived Brian 2.1 to be easy to use and had intent to use the robot again. The ranking of the robot characteristics which the participants stated they liked is presented in Table 17. The most-liked characteristic was identified to be Brian 2.1's human-like voice.

Table 16 Descriptive Statistics for the Adapted Almere Model.

Construct	Minimum	Maximum	Mean	Standard Deviation
PENJ	3	5	4.00	0.53
SP (alpha = 0.56)	1	5	2.87	1.12
TR (alpha = 0.8)	2	5	3.50	0.94
ATT (alpha = 0.6)	3	5	4.13	0.50
PS (alpha = 0.9)	2	5	3.38	1.07
PEOU	2	5	3.50	1.07
ITU	2	4	3.63	0.74
PU (alpha = 0.0)	1	4	2.93	1.00
PAD	3	4	3.63	0.52

Table 17 Most-Liked Robot Characteristics.

Robot Characteristic	Number of Participants That Liked Robot Characteristic
1. Human-like voice	7
2. Providing companionship	6
3. Expressing emotions through vocal intonation and facial expressions	5
4. Life-like appearance and demeanor	4

Table 18 summarizes the participants' prior experiences with computers. The participants of the meal-activity experiments had varying computer skills ranging from no experience to advanced, with the majority of them (i.e., 6 participants) being at least at the beginner's level.

Table 18 Prior Experience with Computers.

Experience Level	Number of Participants
No experience	2
Beginner (email, use simple programs)	2
Intermediate (internet, chat)	2
Advanced (editing documents, using complex programs)	2

Spearman's ρ was used to determine if a correlation exists between computer experience and the participants' ratings of PEOU. A ρ of -0.173 was identified, which showed no significant correlation between these two factors for the participants for an α=0.05. Namely, it was found that all the participants, regardless of their computer experience, found the robot easy to use.

6.3 Discussions

The results from the experiments show that high percentages of participants in both the public and one-on-one user study settings were engaged in the activity interactions with Brian 2.1. This was represented in both interaction scenarios, where participants had visual focus of attention towards both the robot and the activity as well as manipulated the activity objects, i.e., picture cards and meal items. With respect to utterances stated towards Brian 2.1, 15% of the participants playing the memory card game spoke directly to the robot whereas all the participants of the meal-eating activity spoke to the robot. These results may be due to the user study settings themselves, namely, in the public setting it was observed that participants would more readily initiate conversations regarding the interactions with other passersby which included both older adults and caregivers, while in the one-on-one meal interactions, as the participants were alone with Brian 2.1, they directly spoke to it. Both sets of participants had positive responses to the robot's emotional assistive behaviours. This included participants smiling in response to the robot's happy facial expressions and laughing at the robot's jokes. It was interesting to note that even though a large number of participants did not talk to Brian 2.1 in the public setting, a large number (82.5%) did respond to the robot's expressions and jokes. Similar high engagement results have also been observed involving elderly adults in long-term care facilities and animal-like robots including the cat-like robot NeCoRo [173] and the seal-like Paro robot [174]. In both studies, participants would touch and speak to these robots.

A majority of participants in both studies complied with Brian 2.1's instructions, encouragement and help behaviours during the interactions. Compliance in the two settings was 87.5%. It was observed that with respect to the meal-eating activity, compliance was also dependent on users' food preferences, as was evident with one participant who did not like one of the food items on her tray. Even though when Brian 2.1 requested that she eat this item, she would eat another food item on her tray. For such pertinent ADLs it is important that the robot be able to encourage users to eat the high nutrition food items in their meals. This observed scenario is still very much an issue with human caregivers and not just Brian 2.1. Overall, the compliance results show that the elderly participants were willing to follow the robot's prompts to complete the respective ADLs. Only one other study by Fasola and Mataric, [175], has investigated compliance with a human-like robot in such assistive scenarios. Namely, in their study the child-like Bandit II robot instructed an elderly individual through seated physical exercise routines utilizing encouraging speech and demonstrative gestures. They found a high level of participant compliance with the exercise instructions given by Bandit II.

The questionnaire results found that, in general, the participants in both user studies enjoyed interacting with Brian 2.1 and had positive attitudes towards. These results are consistent with the emotional responses from both sets of participants smiling and laughing in response to the robot's emotional behaviours during interactions. The participants also indicated that they trust Brian 2.1 which is consistent with the high compliance rates obtained with respect to the robot's prompts for both activities. As the participants in both studies interacted with Brian 2.1 through natural human communication modalities, this is most likely why they also rated the robot as easy to use. The participants evaluated Brian 2.1 as being social and adaptive due to its ability to use natural human communication modalities, display different emotions and provide encouraging behaviours based on the activity task at hand. The aforementioned positive results are consistent with the participants in both studies wanting to use Brian 2.1 again in the future. The participants of the memory card game scenario also rated Brian 2.1 high for social presence and perceived usefulness. For the memory game scenario, Brian 2.1 displayed all its emotions including sadness. However, for the meal-eating scenario, the robot never displayed a sad emotion due to the fact that the participants were always engaged in the interaction and did not at any time become disengaged. By displaying more emotions during the memory game scenario, the participants during this setting may have attributed a higher social presence to Brian 2.1. This is also consistent with the most-liked characteristic results for Brian 2.1 for the memory game scenario being its ability to express emotions through vocal intonation and facial expressions. With respect to perceived usefulness, it is postulated that since the one-on-one meal-eating interactions were done in a separate room rather than in the dining hall, where other individuals would also be eating their meals, the future intended use of Brian 2.1 was not as clear as in the memory game scenario. Additionally, there was no significant relationship found between the participants' perceived ease of use of Brian 2.1 and their experience with computers for both interaction scenarios, indicating that the participants of both experiments found the robot easy to use regardless of their computer experience. Overall, both questionnaire results identified that the elderly participants liked the fact that Brian 2.1 was able to utilize natural human communication modalities during the two assistive interactions.

6.4 Summary

In summary, this chapter presented two user studies conducted with Brian 2.1 to investigate how elderly users interact with the robot in two different interaction scenarios. The first user study involved placing Brian 2.1 in a public setting at a local long-term care facility where elderly participants interacted with it during a cognitively stimulating memory game. The second user study involved one-on-one interactions between elderly participants and Brian 2.1 during the important meal-eating activity. The results of both user studies showed that the majority of elderly participants were engaged in the activities with Brian 2.1 and complied with the robot's behaviours. Post-study questionnaire results indicated that the elderly participants from both studies enjoyed interacting with Brian 2.1 and found the robot easy to use due to its natural human communication modalities including speech, facial expressions, gestures and vocal intonation. Overall, these results show the potential of using the socially assistive robot Brian 2.1

to assist the elderly to complete tasks in both an open public setting and a one-on-one private setting.

Chapter 7
Detecting Affect from Static Body Language Experiments

This chapter investigate the performance of the robot integrated automated affect from static body language system in being able to recognize and classify a person's accessibility levels during HRI. The first set of HRI experiments investigated implementing the proposed system with two different sensory systems. The second set of |HRI experiments investigated the performance of the proposed system in comparison to a commercially available system.

7.1 Experiment 1: Testing Sensory Systems

Two social HRI experiments were conducted in a laboratory setting to investigate the performance of the affective static body language recognition and classification methodology with each of the sensory systems discussed in Chapter 3: i) a thermal camera and TOF camera, and ii) the 2D camera and 3D depth imager of the Kinect™ sensor. These experiments were also performed to validate the proposed system's robustness to different imaging sensors that can be used by a robot during HRI.

7.1.1 Methods and Participants

Twenty students, ages 19 to 35, participated in the experiments, each naturally implementing a number of different static body poses during the interactions for detection, identification, and categorization into accessibility levels towards the robot. If participants were wearing long sleeves, they were asked to roll up their sleeves to their elbows. 3D and 2D or thermal sensory information was analyzed using the automated identification and classification technique. The participants were split up into two groups with respect to the two sets of experiments.

Both experiments consisted of having the participants interact with Brian 2.1 during a 4 stage interaction scenario. 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 stage was approximately 5 minutes long and the robot displayed a number of different body gestures and facial expressions throughout the interactions. The four stages of the interaction included: i) an Introduction Stage - where Brian 2.1 introduced himself and his capabilities and asked introductory questions of the person with whom it was interacting, such as “What is your name?”, during this stage Brian 2.1 would display body gestures such as pointing and waving as well as displaying happy and neutral facial expressions; ii) a Story Telling Stage - where Brian 2.1 told a story about its experience in a park to the participant, while telling the story, the robot displayed gestures and facial expressions corresponding to the mood in the story; iii) a Memory Stage – where Brian 2.1 would ask a participant to remember items it spoke about in the interaction, and then near the end of the interaction asked the participant to recall the items, the robot would be happy if a participant provided the correct response and sad if the participant did not; iv) the Repetitive Stage - Brian 2.1 would repeat a specific behaviour over and over again, i.e., asking the person the same question continuously using a neutral facial expression and an arms crossed body pose.

7.1.2 Results

In general, approximately 150 different poses were analyzed utilizing the proposed body language recognition and classification technique. The average computation time for the body part segmentation, static pose identification, ellipsoid model generation and body pose classification modules was 55 ms on a Dell workstation with Intel Zeon 3.4 GHz CPU and 12 Gb RAM. Experimental results for both sets of experiments are presented in Figure 20 and Figure 21. It should be noted that 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. In particular, Figure 20 shows ten typical body poses of participants interacting with the robot as captured by the thermal camera and TOF camera sensory system. The thermal and 3D information provided from the sensory system as well as the corresponding ellipsoid and the accessibility levels determined from the proposed technique are presented. In particular, Figure 20 includes the following body poses: (a) upper and lower trunks in neutral positions (with respect to the robot) with the hands resting on the hips, (b) upper and lower trunks in a towards positions with the arms alongside the trunks, (c) upper and lower trunks in an away position with the arms alongside the trunks, (d) leaning forward in a towards position with arms behind the trunks, (e) upper and lower trunks in neutral positions with the elbow of one arm resting on the other arm, (f) upper and lower trunks in a neutral position with one arm up and the chin resting on that arm, (g) upper and lower trunks in a towards position with arms crossed in front of the trunks, (h) upper and lower trunks in a neutral position with the hands intertwined behind the head, (i) upper and lower trunks in a towards position with hands clasped in front of the trunks, (j) upper and lower trunks in a toward position while leaning forward and hands clasped in front of the trunks. Figure 21 presents ten typical static body poses displayed by the participants during the interaction stages with the robot utilizing the Kinect™ sensory system. The 2D and 3D sensory information, ellipsoid models and accessibility levels for each pose are presented in the figure. In particular, the following body poses are shown: (a) and (b) poses with the upper and lower trunks in towards position with the arms at the sides of the body, (c) upper and lower trunks in towards position with one arm up and the chin resting on that arm, (d) upper trunk in a neutral position and the lower trunk in a towards position, (e) upper and lower trunks in the towards position while leaning forward with hands on the hips, (f) upper and lower trunks in towards position with one hand grasping the elbow of the other arm, (g) upper and lower trunks in towards position with the arms behind the back, (h) upper and lower trunks in neutral position with arms at the sides, (i) upper and lower trunks in neutral position with arms crossed, and (j) both upper and lower trunks in away position with arms at the sides.

7.1.3 Accessibility Baseline Coding

An expert coder was used to determine the baseline accessibility levels for all the poses displayed in the experiments. In particular, 2D images of the participant's poses from both sensory systems were provided to the expert coder to classify the appropriate accessibility levels of the poses. For the set of experiments utilizing the thermal and TOF camera-based sensory system, the percentage of agreement between the accessibility levels determined by the automated system and the baseline levels was determined to be 80%. For the Kinect™ sensory system, the percentage of agreement between the automated system and the baseline levels was 84%. Hence, the overall percentage of agreement for the experiments was found to be 82%. These high agreement rates validate the use of the proposed body language identification and categorization technique for HRI scenarios.

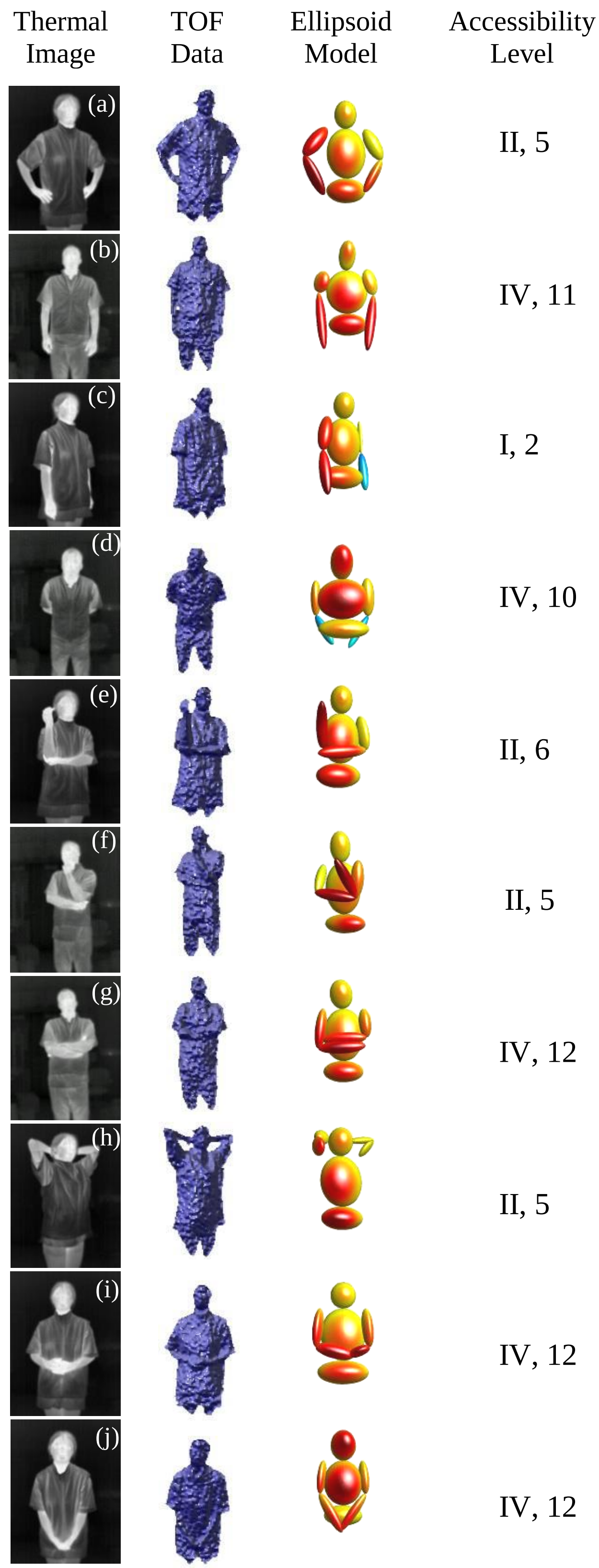


Figure 20 Experimental Results with Thermal and TOF cameras. [121]

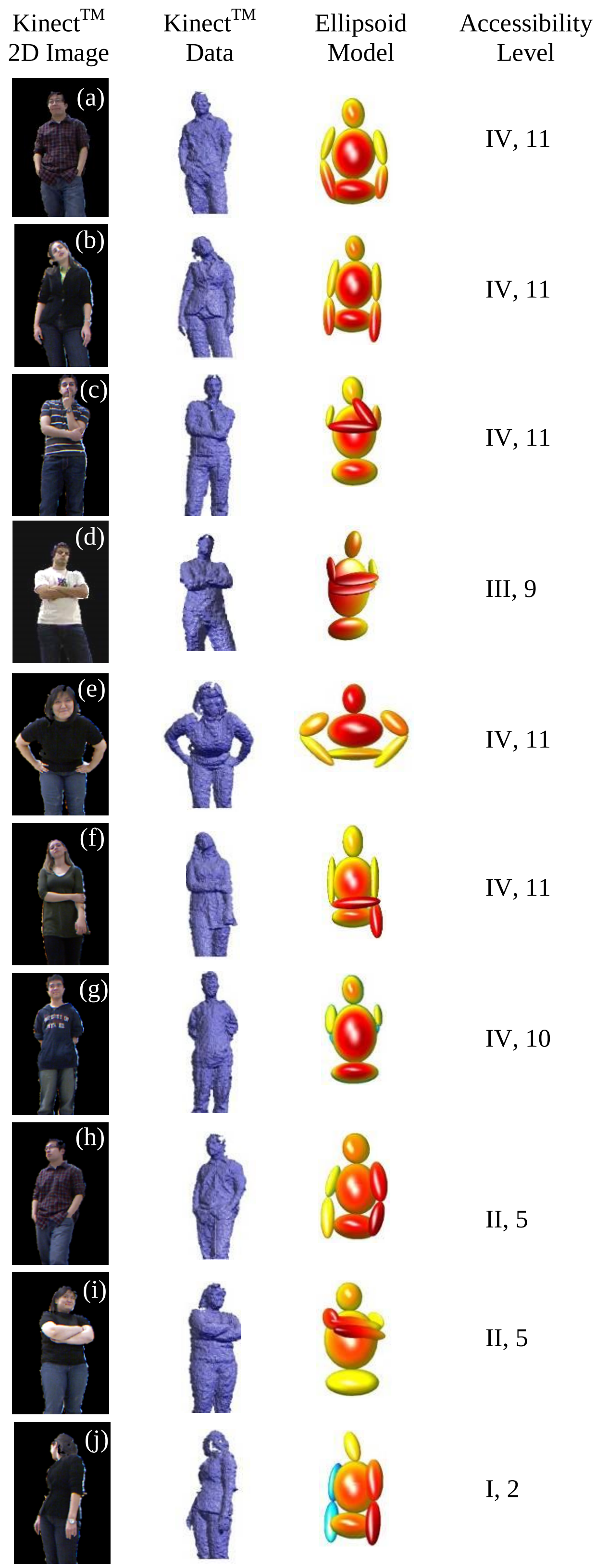


Figure 21 Experimental Results with Kinect™ Sensor. [121]

7.1.4 Discussion

The ellipsoid models generated utilizing information from both sensory systems provide a close approximation of the poses displayed by the participants during the interactions. This can be seen by comparing the participant's poses in the sensory information to the ellipsoid models in both Figure 20 and Figure 21. The results also show that the proposed affect recognition technique is independent of the sensory system used to obtain the input data and hence, can be implemented with different onboard sensory systems. The results of very similar poses obtained from both sensory systems can be seen, for example, in Figure 20(c) and Figure 21(j), Figure 20(g) and Figure 21(i), and Figure 20(f) and Figure 21(c). Hence, the proposed approach is not affected by large differences in sensory resolution, in particular, the 3D data of the TOF camera has a resolution of only 176 x 144 pixels while the 3D images from the Kinect^{TM} sensor have a resolution of 640 x 480 pixels. This difference in 3D resolution is evident in the 3D data presented in Figure 20 and Figure 21.

Examples of poses that have occluded body parts can be seen in Figure 20(c) and (d), and Figure 21(g) and (j). Body parts are found to be occluded when the corresponding skin region or 3D data for that particular body part is not present in the sensory data due to the static pose that is displayed. The current approach estimates the location of occluded body parts utilizing the shape and pose of the ellipsoid for that particular body part from previous frames. Namely, if a body part is found to be occluded, the ellipsoid from the previous frame is connected to the adjoining body part of the current frame utilizing the skeleton model. If both the lower arm and upper arm of a single arm are occluded, the ellipsoids from the non-occluded arm are positioned symmetrical to represent the occluded arm. It should be noted that this estimation technique does not affect the classification of the accessibility level of the person and therefore is sufficient for this work.

From the classification results from both experiments, it was found that 61% of the body poses observed during all the interactions were classified as accessibility level IV, 2% as level III, 27% as level II, and 10% as level I. The most common finer-scaling level observed was level 11 with a rate of 31% with respect to all the poses displayed. From these results, it can be seen that the participants were in accessibility level IV the majority of the time, and hence, were both open to interactions with the robot as well as stayed engaged during the interactions. The low number of level III body poses observed is partly due to the fact that the participants more often rotated their upper and lower trunks together for the N and T orientations rather than separately, where the latter mainly results in an accessibility level III classification. It was also observed that the majority of the participants were in accessibility levels I and II during the Story Telling and Repetitive stages, in which the interactions were more one-directional as the robot implemented its behaviours without any input from the participants. The classification results that did not agree with the expert coder were mainly for the poses where other body parts such as the arm(s) occluded a large part of the trunks. In such scenarios, the sensory systems had less 3D data to estimate the orientation of the trunks. The slightly higher percentage of agreement with the Kinect™ sensory system was due to the fact that there was less occlusion of the trunks during the interactions.

7.2 Experiment 2: Performance Comparison

A performance comparison is presented by comparing the identified static body poses obtained from the proposed approach using the Kinect™ 2D and 3D depth images to poses identified using the human body skeleton joints obtained from the Kinect™ SDK [62]. Furthermore, the accessibility levels obtained from the 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.

7.2.1 Methods and Participants

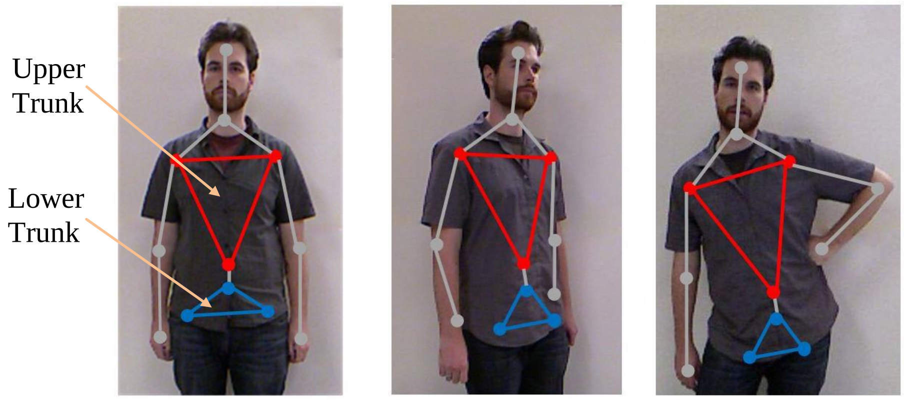
Eighteen participants, aged 19 to 35 (μ = 24, σ = 5.33) participated in the performance comparison study. Similar to the first set of experiments, Brian 2.1 was operated in a Wizard of Oz fashion. Each participant interacted with the robot in four different interaction stages: 1) Introduction Stage, where the robot would introduce itself to the participant, 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 behaviour for 5 minutes. 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 towards Brian 2.1.

7.2.2 Kinect^{TM} SDK Body Pose Estimation Approach

The Kinect^{TM} SDK utilizes a random decision forest and local mode finding to generate joint locations of up to two people from depth images [176]. The person closest to the robot is identified as the user. A technique to identify the static body pose orientations and leans was developed in order to determine accessibility levels from the Kinect^{TM} SDK joint locations during social HRI.

To determine the orientations of the upper and lower trunks in a static body pose utilizing the joints provided by the Kinect^{TM} SDK, the joint positions of the left and right shoulders as well as the spine point (in the middle of the trunk along the back) are utilized to define the upper trunk, whereas the left, right, and hip center joint points are utilized to define the lower trunk. Each set of three points are used to define a plane in 3D space representing either the upper or lower trunk, Figure 22(a). The relative angle between the normal of each plane and the Kinect^{TM} camera axis is then used to determine a person's upper and lower trunk orientations with respect to the robot.

The trunk lean is determined by identifying the relative position of the shoulders with respective to the hips. Specifically, if the left shoulder is greater than 30% of the hip width (distance between left and right hip joint positions) to the left of the left hip joint, then a person is in a left sideways lean. A similar procedure is used to define a right sideways lean. Forward leans are defined as a trunk configuration that results in an angle greater than 10^o between the normals of the planes defined for the upper and lower trunks. These parameters have been verified through numerous testing with different individuals having various body shapes and sizes. Example trunk orientations and leans are shown in Figure 22. In addition to the upper and lower trunks, the head is defined by the joints at the top of the head and shoulder center, and the upper arms are defined by the shoulder and elbow joints. The left and right lower arms are defined by the left and right elbow and hand joints.



(a) Upper and Lower Trunks in T Orientations

(b) Upper and Lower Trunks in A Orientations

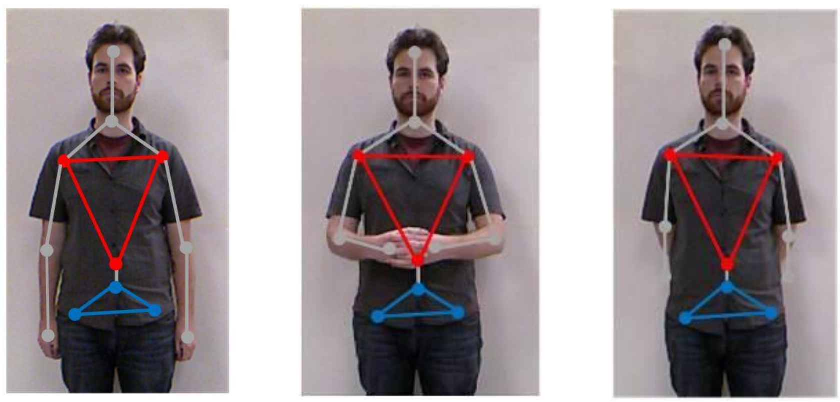
(c) Right Sideways Lean

Figure 22 Example trunk orientations/leans using joint locations provided by the Kinect™ body pose estimation technique.

The arm orientations are determined by the relative distances between the lower arms and the upper trunk. The distances of each lower arm are defined as z_l and z_r, which are the depth distances from the centroid of the left and right lower arms to the robot, respectively. The average of the left and right lower arm distances, z_a, is compared to the distances of the upper trunk joints to determine a person's arm orientation. The upper trunk joint distances are defined as $z_{leftshoulder}$, $z_{rightshoulder}$, and z_{spine}, the depth distances of the left shoulder, right shoulder and spine joints to the robot, respectively. The arm orientation classifications of T, A and N utilizing these distances are shown in Table 19. Example N, T and A arm orientations are shown in Figure 23. Once the trunk and arm configurations are determined, NISA is utilized to identify a person's degree of accessibility with respect to the robot.

Table 19 Arm Orientation Classification for Kinect™ Body Pose Estimation Approach.

Arms in Toward (T) Orientation	
$z_a < z_{tmin}$, where $z_{t\min} = \min(z_{leftshoulder}, z_{rightshoulder}, z_{spine})$ and $z_a = (z_l + z_r)/2$	(20)
Arms in Away (A) Orientation	
$z_a > z_{tmax}$, where $z_{t\max} = \max(z_{leftshoulder}, z_{rightshoulder}, z_{spine})$	(21)
Arms in Neutral (N) Orientation	
$z_{tmax} < z_a < z_{tmin}$	(22)



(a) Arms in N Orientation (b) Arms in T Orientation (c) Arms in A Orientation

Figure 23 Example arm orientations using joint locations provided by the Kinect™ body pose estimation technique.

7.2.3 Accessibility Baseline Coding

To investigate the reliability of the proposed automated affect from static body language system and Kinect™ 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 2D image from the Kinect™ 2D 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.

7.2.4 Results and Discussion

Overall, the participants displayed 223 different static poses during the experiments. Figure 24 shows eight representative poses displayed during the aforementioned interaction experiments with Brian 2.1. The first and second columns present the 2D color and 3D data of the segmented static poses. The corresponding ellipsoid models obtained from the multimodal pose estimation are shown in the third column. Lastly, the multi-joint models obtained from the Kinect™ SDK body pose estimation approach are presented in the fourth column.

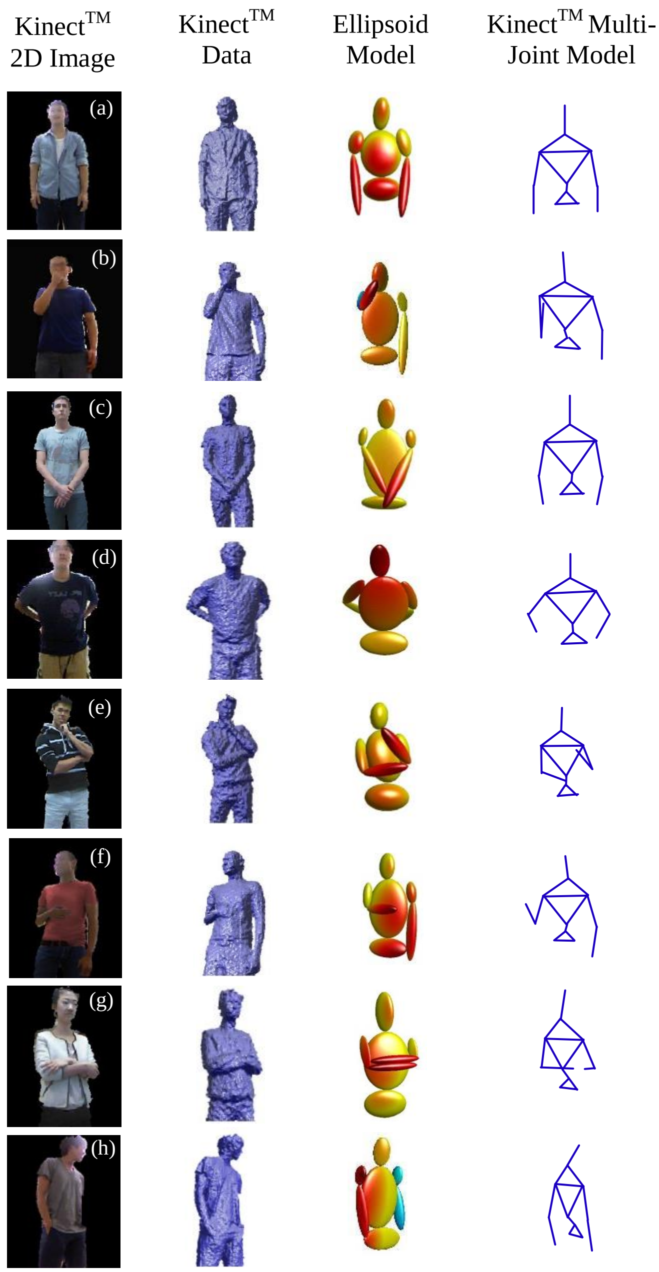


Figure 24 Example static poses. [121]

The typical poses displayed consisted of the following: (a) arms at the sides and both trunks in toward positions, (b) one hand touching the head and both trunks in toward positions, (c) hands touching in front of the trunk with the upper trunk in a toward position and the lower trunk in a neutral position, (d) leaning forward and arms away with both trunks in toward positions, (e) 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 lower trunk in a toward position, (f) one arm crossing the trunk with the other at the side with both trunks in neutral positions, (g) arms crossed in front of the trunk with both trunks in neutral positions, and (h) 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 20.

Table 20 Accessibility Level Results.

	Multimodal Static Body Pose Estimation Approach		Kinect™ Body Pose Estimation Technique	
Body Pose	Accessibility Level	Finer-Scaling	Accessibility Level	Finer-Scaling
Figure 24(a)	IV	11	IV	11
Figure 24(b)	IV	12	II	6
Figure 24(c)	IV	12	IV	11
Figure 24(d)	IV	10	IV	10
Figure 24(e)	III	9	II	6
Figure 24(f)	II	6	II	5
Figure 24(g)	II	6	II	6
Figure 24(h)	I	2	I	2

Similar to the previous experiments, the ellipsoid body models of the multimodal static body pose estimation approach created using the 3D and 2D 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 2D images and 3D depth information in Figure 24. It should be noted that the skin color of the participants did not influence body part segmentation or ellipsoid model generation. Again, blue ellipsoids indicate occluded body parts in Figure 24. In Figure 24(b), the blue ellipsoid represents an upper arm that is directly behind a lower arm and in Figure 24(h) the blue ellipsoids represent the upper and lower right arms. 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.

It can be seen from Figure 24 that with certain body poses, the Kinect™ SDK body pose estimation does not accurately identify the correct poses of the arms, in particular when a participant's arms are touching each other or other body parts. For example, in the Kinect™ 3D multi-joint body model of Figure 24(b) the hand of the multi-joint model does not make contact with the face or head. For the pose displayed in Figure 24(c) the hands are not clasped in the multi-joint model. In Figure 24(e) the right arm is not touching the left arm and the left arm is not touching the head. In Figure 24(f) the participant's right arm is not crossing in front of the trunk. Finally, in Figure 24(g) the arms are not crossed. The random 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 [176]. 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 [176]. Hence, it is postulated 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 3D data, allowing for multiple initial poses.

7.2.5 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™ 3D multi-joint body model, Table 21. The 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^{TM} 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^{TM} body pose estimation approach which has a fair strength of agreement [133].

Table 21 Performance Comparison Statistics.

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

7.3 Summary

In summary, this chapter presented two sets of unique experiments to validate the performance of the proposed approach. The first set of experiments, investigating two different imaging-based sensory systems has shown the robustness of the affective static body language classification technique to different robotic sensors. The second set of experiments consisted of a performance comparison study which showed that the proposed 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^{TM} SDK joint locations. These experiments also validate the integration of the body language recognition and affect classification sensory system into a social robot such as Brian 2.1 in order for the robot to recognize affective body language during one-on-one interactions.

Chapter 8
Detecting Affect from Dynamic Body Language Experiments

This chapter investigates the performance of the proposed affective dynamic body language estimation system. The affective dynamic body language estimate system is applied to the data obtained from the HRI experiments conducted in Chapter 6 at the local long-term care facility between Brian 2.1 and elderly adults during the meal-eating activity.

8.1 Methods

With respect to detection and identification system, each participant sat at a table across from Brian 2.1, at a distance between 1.2-2.0m from the robot. An example HRI scenario is shown in Figure 25. Both 2D and 3D information from the Kinect™ sensor mounted on Brian 2.1's chest were used to record the participant's natural displays of body language during the meal activity. 3D data was obtained every 10th frame from the 60 Hz Kinect™ sensor for upper body tracking to obtain the body features needed for automated affective body language classification.

In order to identify affective upper body language, all body movements and postures except for those directly involved with performing meal actions were identified as body language displays. Meal actions are defined as body movements that result in the manipulation of the meal items, such as moving the utensil to consume food or lifting the beverage cup. These actions were identified using the utensil tracking system and the meal tray embedded sensors.

The skeleton model for each body language display is utilized to calculate the body language features for that display. The body language features are then classified into both valence and arousal levels using the learning techniques described in Chapter 4 via the Weka Data Mining Software [130]. The corresponding 2D video data from the interactions were used for human baseline coding of each participant's affective body language.

Figure 25 Example HRI scenario.

8.2 Affective Body Language Baseline Coding

The performance of each automated affect classification learning technique is investigated by determining the agreement between the classification results and observer baseline coding. Videos of the participants interacting with Brian 2.1 during the meal activity were presented to 21 observers (3 females and 18 males). These observers ranged in age from 18 to 32 (μ = 22.9, σ = 3.5). Each observer was asked to identify the affective state displayed by the body language of the elderly participants. The faces of the participants were covered in the videos and audio was removed to ensure observers were identifying affect from body language alone. The average total length of a single body language display video was three minutes with an average of approximately nine different body language displays. The videos were presented to each observer in a random order. Example frames from the videos of two participants are presented in Figure 26. The observers coded both valence and arousal levels for each body language display using a five point Likert scale, i.e., valence: 1 = very unpleasant, 2 = unpleasant, 3 = neutral, 4 = pleasant, 5 = very pleasant; and arousal: 1 = very inactive, 2 = inactive, 3 = moderately active, 4 = active, 5 = very active. The median value for the observer codings was utilized as the baseline label for each body language display, similar to the approach in [101]. Inter-coder reliability for both valence and arousal was determined using Cronbach's alpha. Alpha values of 0.71 and 0.72 were obtained for valence and arousal, respectively, which are defined to represent acceptable agreement between the observers. The results of the coding identified arousal levels ranging from inactive to moderately active, and valence levels ranging from unpleasant to very pleasant.

(a) Participant bowing her trunk towards the robot

(b) Participant closing and opening her arms

Figure 26 Example video frames obtained from Kinect^{TM} 2D camera.

8.3 Affect Estimation Results and Discussion

10-fold cross-validation [177] was performed to test each of the learning-based classification techniques. Table 22 presents the recognition rates for both valence and arousal for the classification techniques. The mean-absolute-error (MAE) for each technique is also presented in Figure 27.

Table 22 Cross-Validation Results for the Classification Techniques.

Classification Technique	**Valence Recognition Rate (%)**	**Arousal Recognition Rate (%)**
Naïve Bayes	70.7	92.1
Logistic Regression	70.0	90.7
Random Forest	72.9	93.6
KNN	72.1	89.3
Adaboost with Naïve Bayes	70.0	93.6
RBF Network	77.9	91.4
SVM	66.4	83.6

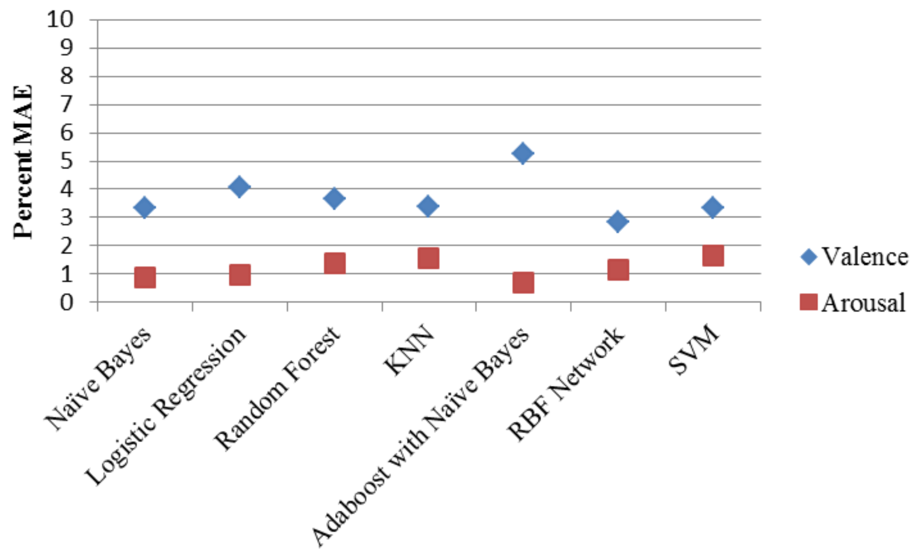


Figure 27 Mean-absolute-errors for the classification techniques.

The Random Forest and Adaboost techniques both had the highest arousal recognition rate of 93.6% and the RBF Network had the highest valence recognition rate of 77.9%, while SVM had the lowest recognition rates for both valence (66.4%) and arousal (83.6%). The Cochran's Q test was utilized to determine that there were statistically significant differences between the recognition rates of the classifiers for both arousal (Q = 52.8, $p < 0.001$) and valence (Q = 51.3, $p < 0.001$) levels. Post-hoc McNemar tests with Bonferroni correction ($p < 0.00238$) were then performed to determine if the aforementioned recognition rates resulted in statistically significant differences between the classifiers. Even though the Random Forest and Adaboost techniques had the highest recognition rates for arousal, their recognition rates were only statistically significantly higher than the SVM technique ($p < 0.001$). Furthermore, SVM was found to have a significantly lower recognition rate for arousal when compared to all the other techniques ($p < 0.002$) except for KNN ($p = 0.008$). Although the RBF Network technique had the highest recognition rate for valence, it was found to only have a statistically significantly higher recognition rate than the Naïve Bayes ($p = 0.002$), Logistic Regression ($p = 0.001$), Adaboost ($p = 0.001$) and SVM ($p < 0.001$) classifiers. It is interesting to note that boosting the Naïve Bayes classifier did not significantly improve the recognition rates for valence or arousal. In general, the affect classification techniques were able to classify arousal with higher recognition rates than valence. This is consistent with other studies that have investigated the recognition of arousal and valence with respect to other types of body movements [98].

The classification results presented herein show that natural affective body language of seated older adults during assistive HRI can be classified using standard learning techniques. They also motivate the integration of a learning-based dynamic body language classification system onto the human-like social robot Brian 2.1 in order for the robot to recognize and classify natural seated body language during one-on-one interactions using non-contact sensing techniques.

Further investigation into the arousal levels identified during body language classification found that all the participants had an average moderately active arousal level for at least one meal. One participant had an average inactive arousal level for both meals. Through observations of video recordings of the interactions, it was found that this participant had difficulty eating her food items and required more prompts than the others but still interacted with the robot by smiling and laughing at its jokes and speaking to it, e.g., “It is very nice to be here with you.” With respect to valence, 5 out of 8 participants had an average pleasant valence level for one meal, while the three other participants had an average neutral valence level. For the other meal, 7 participants had an average neutral valence level and one participant had an average unpleasant valence level. Two participants had instances of body language displays corresponding to very pleasant valence levels during both meals. No very unpleasant valence levels were identified. The aforementioned participant that had difficulty eating her meal also had instances of body language displays identified as unpleasant. The detection of this range of affective states motivates the monitoring of affective body language by Brian 2.1 in order for the robot to determine its appropriate assistive behaviours during the meal-eating activity. Namely, monitoring arousal and valence levels of a user will allow Brian 2.1 to effectively adapt its behaviours in response to the user’s affective state in order to encourage user engagement in the meal activity, independent eating habits, and social interactions during the activity.

8.4 Summary

In summary this chapter investigated the ability of the proposed automated affective dynamic body language recognition and classification system to interpret natural seated and dynamic displays of affective human body language during assistive HRI with the elderly. The proposed system utilizes 3D data from an onboard Kinect™ sensor to recognize and track body language features using the 3D skeleton of a person’s upper body. The features are then classified into valence and arousal values using learning-based classifiers. A variety of different potential classifiers have been investigated and compared for the proposed application. Classification results obtained from assistive HRI experiments with older adults and Brian 2.1 during the important meal-eating activity have shown that natural dynamic affective body language of seated older adults can be recognized and classified using learning techniques.

Chapter 9
Affect-Aware Robot Behaviour Experiments

This chapter presents an experiment that uniquely investigates how people actually interact with a socially assistive robot which explicitly uses identified accessibility levels throughout social interaction to determine its own behaviours. To do this two robot behaviour types are compared to determine if an accessibility-aware emotionally responding robot influences the overall interactions with individuals: 1) the robot determines its assistive behaviours based on the state of the activity, herein defined as the non-accessibility-aware behaviour type, and 2) the robot determines its assistive behaviours based on the accessibility level of the person as well as the state of the activity, herein defined as the accessibility-aware behaviour type. The proposed affective static body language recognition and classification system with the Kinect™ sensory was utilized during the HRI experiments. These experiments were conducted using two assistive scenarios: 1) Robot Tutor, and 2) Robot Restaurant Finder. The Robot Tutor scenario consisted of Brian 2.1 engaging a participant in memory and logic games. In the Robot Restaurant Finder scenario, the robot assisted a person in choosing a restaurant to go to for dinner.

9.1 Participants

Twenty-four participants ranging in age from 20−35 (μ = 24.7, σ = 4.4) participated in the study. Each participant interacted with the robot twice, once with each behaviour type with one week between interactions. During each interaction the robot 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 non-accessibility-aware robot first. The order of assistive scenarios (Robot Tutor or Robot Restaurant Finder) was also counterbalanced.

9.2 Interaction Scenarios

9.2.1 Robot Tutor Interaction

The Robot Tutor 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 stage was designed to test a participant's ability to extract meaning from complex information. Three logic questions were asked by the robot. 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 Robot Tutor 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 would finish with Brian 2.1 informing the user that all the games were finished.

9.2.2 Robot Restaurant Finder Interaction

The Robot Restaurant Finder interaction consisted of the robot assisting a participant in identifying and locating a suitable restaurant for dinner based on a participant's preference. This interaction had four main stages: 1) greeting, 2) information gathering, 3) restaurant suggestion, and 4) providing directions. Similar to the Robot Tutor 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 preference 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 from the current location.

9.3 Robot Behaviour Design

The robot's behaviour types were implemented using finite state machines (FSMs). The input into the FSM from the participant for the robot's non-accessibility-aware behaviour was speech. Whereas, speech and accessibility levels were both used as inputs into the FSM for the robot's accessibility-aware behaviour. An operator was utilized 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., [178, 179]. 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 behaviour types determined the robot's behaviour based on the current state of the interaction scenario and the corresponding inputs from the participants.

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

9.3.1 Brian 2.1's Non-Accessibility-Aware Behaviours

With respect to Brian 2.1's non-accessibility-aware behaviours, the robot replies to positive responses during the Robot Tutor and the Robot Restaurant Finder interactions by verbally acknowledging the responses. For example, during the Robot Tutor interaction one reply to a positive response is "Yes that answer is correct." During the Robot Restaurant Finder 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 [180] in order for the participant to identify a positive response. This is achieved by giving an example answer during the Robot Tutor interaction or by restating the question during the Robot Restaurant Finder interaction. To re-engage a participant who did not respond to the robot, for both the Robot Tutor and Robot Restaurant Finder interactions, Brian 2.1 asks the participant if he/she would like it to repeat its previous statement. During the interactions, the behaviours 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 non-accessibility-aware behaviours are summarized in Table 23. Figure 28 shows a visual example of this robot behaviour type.

Table 23 Non-Accessibility-Aware Robot Behaviours.

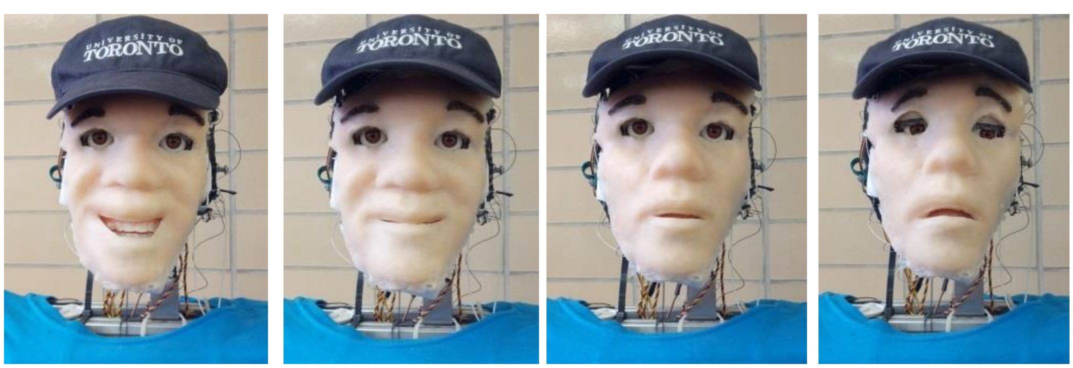
User Speech Input	Robot Behaviour	Example Speech during Tutor Interaction	Example Speech during Restaurant Finder Interaction
Positive Response	Verbally acknowledges positive response.	"Yes that answer is correct."	"Okay. Next, where would you prefer to eat? Close to here, downtown or in the entertainment district?"
Negative Response	Provides verbal assistance.	"Unfortunately, that is incorrect. Forty-two divided by 7 is 6. What is the answer to 6×11?"	"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."	"I can repeat the question for you."

Figure 28 Brian 2.1 providing verbal assistance while swaying its trunk during the non-accessibility-aware behaviour type.

9.3.2 Brian 2.1's Accessibility-Aware Behaviours

The goal of the accessibility-aware interactions is to detect a person's accessibility levels towards Brian 2.1 and utilize emotional robot behaviours 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 behaviours communicates empathy towards the participant [181], which is important for building rapport and trust between communicators [182]. Emotional behaviours displayed by the robot during social interaction can also improve user engagement [183] and affect [184, 185] as well as encourage correct responses during learning scenarios [186]. 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 non-accessibility-aware behaviours. An open mouth smile is used as it distinguishably conveys increased positive valence as compared to a closed mouth smile [187]. The robot displays positive valence using a closed mouth smile and a happy tone of voice. Neutral valence is simply displayed utilizing 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 Figure 29.



(a) high positive valence (b) positive valence (c) neutral valence (d) negative valence

Figure 29 Brian 2.1's facial expressions.

When a participant is in accessibility level IV, during the Robot Tutor interaction, the robot encourages a positive (correct) response by verbally congratulating the participant, enthusiastically nodding its head and clapping its hands while displaying high positive valence, Figure 30(a). Such behaviours have all been shown to convey positive emotions [188, 189] and positively reinforce desired behaviours in others [190]. During the Robot Restaurant Finder interaction, Brian 2.1 verbally acknowledges a positive response with high positive valence while nodding its head enthusiastically, Figure 30(b). For a negative response, during both interaction 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 behaviours with respect to a person's positive, negative or no response behaviour. For participant behaviours displayed in accessibility level III, the robot displays positive valence without the nodding or clapping gestures. Removing such non-verbal gestures reduces the level of positive reinforcement [191]. When a participant is in accessibility level II, the robot utilizes neutral valence behaviours to neither reinforce nor punish him/her [192]. For a participant in accessibility level I, Brian 2.1 displays negative valence and waves its arm in a beckoning gesture, Figure 30(c). The combination of the beckoning gesture and the sad facial expression is used to get the person's attention [193] and evoke sympathy which motivates a person to help the robot [194, 195].

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. Humor is utilized in addition to the emotional 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 [196]. Telling jokes has also been shown to improve users' task enjoyment during HRI [197]. After delivering the punchline of the joke, the robot lifts its hand to cover its mouth while giggling, as seen in Figure 30(d). At the end of the interaction, Brian 2.1, while displaying positive valence, waves goodbye and thanks the user for participating, Figure 30(e). The robot's accessibility-aware behaviours based on the verbal responses and accessibility levels of participants are presented in Table 24.

(a) Brian 2.1 displaying high positive valence while congratulating a user and clapping

(b) Brian 2.1 displaying high positive valence while acknowledging a positive response and nodding

(c) Brian 2.1 displaying negative valence while offering assistance and using a beckoning gesture

(d) Brian 2.1 displaying positive valence while telling a joke and giggling

(e) Brian 2.1 displaying positive valence while saying goodbye and waving

Figure 30 Example accessibility-aware robot behaviours.

Table 24 Accessibility-Aware Robot Behaviours.

User Input		Robot Behaviour	Example Speech during Tutor Interaction	Example Speech during Restaurant Finder Interaction
Acces-sibility Level	Speech Input			
IV	Positive Response	Acknowledges/Congratulates with high positive valence, head nodding and/or clapping.	“Excellent, can you think of another one?”	“Italian food sounds great!”
	Negative Response	Thanks participant for responding and offers assistance with high positive valence.	“Thank you for your answer. Unfortunately it is not correct. Another example is baseball, but not hockey. Please try again!”	“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.”	“I can repeat the question for you if you would like.”
III	Positive Response	Acknowledges/Congratulates with positive valence.	“Great job! How about we try a different game now.”	“Great choice, the entertainment district has a lot of good restaurants.”
	Negative Response	Thanks participant for responding and offers assistance with positive valence.	“Thanks, that was a good try but your answer is not correct. Here is another example: radiator starts with the letter R and is part of a car.”	“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?”	“Why don’t I repeat the locations I know about?”
II	Positive Response	Acknowledges/ Congratulates with neutral valence.	“Good job. Let’s move on to the next game.”	“Good idea, it’s important to try and save money.”
	Negative Response	Offers assistance with neutral valence.	“That answer is not correct, another example is apples but not oranges.”	“What you said is a sum of money I am not familiar with. How about spending $40 to have a gourmet meal.”
	No Response	Offers to repeat its last statement with neutral valence.	“Why don’t I repeat what I said?”	“Why don’t I repeat the last question?”
I	Positive Response	Acknowledges response with negative valence and a beckoning gesture.	“That answer is right. My turn. How about correct, but not wrong.”	“Sushi it is!”
	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.”	“I am sorry I did not understand your last statement. Do you want me to make reservations for you?”
	No Response	Repeats previous statement with negative valence and a beckoning gesture.	"Excuse me, did you hear the question? How many degrees are there between the hands of a clock pointing to 6:30?"	“Excuse me, did you hear the question? Do you need directions to the restaurant?”

9.3.3 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 Behaviour Questionnaire

(SBQ) [198]. The SBQ was developed specifically to measure user perceptions of a robot's social intelligence with varying types of social behaviours [198]. Cronbach's alpha has determined the reliability of the SBQ to be 0.7-0.9, [198], 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 non-socially intelligent agents [198]. Table 25 shows the constructs and corresponding statements that were used in the questionnaire. Responses to the questionnaire were obtained by each participant indicating his/her agreement with each statement using a 5-point Likert scale (1=strongly disagree, 2=disagree, 3=neutral, 4=agree, 5=strongly agree).

Table 25 Social Behaviour Questionnaire Statements and Constructs.

Constructs	Statements
Altruism	1. The robot makes people feel welcome.
	2. The robot loves to help others.
	3. The robot looks down on others.
Assertiveness	4. The robot automatically takes charge.
	5. The robot tries to lead others.
	6. The robot can talk others into doing things.
Competence	7. The robot is full of ideas.
	8. The robot learns quickly.
	9. The robot comes up with good solutions.
Dutifulness	10. The robot tries to follow the rules.
	11. The robot keeps promises.
	12. The robot does crazy things.
Empathy	13. The robot feels others' emotions.
	14. The robot anticipates the needs of others.
	15. The robot is concerned about others.
Helpfulness	16. The robot has a soft heart.
	17. The robot goes out of its way for others.
	18. The robot will do anything for others.
Modesty	19. The robot acts like it is more important than others.
	20. The robot thinks that it's better than other people.
	21. The robot doesn't like to attract attention.
Responsibility	22. The robot likes to be in the service of others.
	23. The robot takes others' interests into account.
	24. The robot is able to cooperate with others.
Sociability	25. The robot has little to say.
	26. The robot feels comfortable around people.
	27. The robot is skilled at handling social situations.
Sympathy	28. The robot likes to do things for others.
	29. The robot values cooperation over competition.
	30. The robot is not interested in other people's problems.
Trust	31. The robot trusts others.
	32. The robot acknowledges others' accomplishments.
	33. The robot trusts what people say.

9.4 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 behaviour 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 behaviour types to be more socially intelligent than the other.

9.4.1 Accessibility Classification

The most frequent accessibility levels of the participants obtained by the robot during each stage of the interactions for both behaviour types were compared to 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 affective static body language recognition and classification system appropriately matched 75% of the self-reported levels for all the interactions for both behaviour 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. 85% of the self-reported level 2 ratings were matched with NISA accessibility level II from the automated system. 48% of the self-reported level 1 ratings were matched with NISA accessibility level I from the automated system. It should be noted that overall only 7 participants self-reported a small number of their poses as level 1. The poses that were not identified as NISA accessibility 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 towards 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 obtained from the automated system and the openness levels obtained from the self-study report, $z = 0.393$, $p = 0.695$.

9.4.2 Comparison of Robot Behaviour Types

In total, 1,498 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 non-accessibility-aware interactions and 774 poses obtained during the accessibility-aware interactions. Table 26 summarizes the number of static poses identified for each accessibility level and robot behaviour type. For the non-accessibility-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 accessibility-aware 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 minutes with the non-accessibility-aware robot (6 minutes during the Robot Tutor interaction and 5 minutes during Restaurant Finder interaction) and 12 minutes with the accessibility-aware robot (7 minutes during the Robot Tutor interaction and 5 minutes during Restaurant Finder interaction).

Table 26 Participant Accessibility Levels.

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

It was hypothesized that the participants' accessibility levels would be higher during interactions with the accessibility-aware robot than during interactions with the non-accessibility 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 non-accessibility-aware robot, however, when they interacted with the accessibility-aware 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 behaviour types. These results show that, in general, the participants were more accessible towards the social robot when it had the capability to both recognize and respond to their accessibility levels.

9.4.3 Questionnaire Results

A summary of the mean participant ratings for the constructs of the post-interaction questionnaire are presented in Table 27. 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 [199]. All the constructs obtained alpha values of 0.6 or higher except for Dutifulness, which had an alpha value of 0.2 for the non-accessibility-aware robot behaviour type (Table 27). 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 [200].

A Wilcoxon Signed Rank Test was conducted to compare the overall results for the two robot behaviour types. The results showed that the accessibility-aware robot behaviour type was perceived to be significantly more socially intelligent than the non-accessibility-aware robot behaviour type, $z = 4.332$, $p < 0.001$. This result is similar to the study conducted by Ruyter et al. in [198] 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 non-accessibility-aware behaviour type when compared to the accessibility-aware behaviour type. With respect to Competence, the same mean rating may have been obtained, since for both behaviour types the robot had the knowledge to complete the necessary interaction tasks, which is an indicator of competence [190]. Namely, the robot was always able to identify correct or incorrect participant responses to questions during the Robot Tutor interaction and find a restaurant during the Robot Restaurant Finder interaction.

Assertiveness may have been rated a bit lower for the accessibility-aware behaviour type due to it displaying more body movements/gestures during the interactions. In a study presented in [201], it was found that an increased amount of body movements was an indicator of non-assertiveness 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 [190].

Table 27 Mean Questionnaire Construct Results.

Construct	Mean Rating for Non-Accessibility-Aware Behaviour (Cronbach's Alpha)	Mean Rating for Accessibility-Aware Behaviour (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

9.5 Summary

This chapter presented the first automated static body language identification and categorization system implemented on an accessibility-aware robot that can identify and adapt its own behaviours to the accessibility levels of a person during one-on-one social HRI. The HRI experiments presented in this chapter investigated how individuals interact with an accessibility-aware social robot, which determines its own behaviours based on the accessibility levels of a user towards the robot. The results indicated that the participants were more accessible towards an accessibility-aware robot over a non-accessibility aware robot, and perceived the former to be more socially intelligent. Overall, these 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 socially assistive interactions.

Chapter 10
Recognition of Affective Body Language Displayed by a Human-Like Socially Assistive Robot

This chapter presents novel experiments that demonstrate that body movements and postures for human-like robots can communicate certain affective states and hence, should be considered as an important part of interaction on the robot's side. The first objective of the experiments is to determine if non-expert individuals are able to identify emotions from the body language displayed by Brian 2.1, as described in Chapter 5. The second objective of the experiments is to compare how individuals interpret the same emotional body language displayed by the robot and a human actor. Participants were asked to watch videos of both Brian 2.1 and an actor displaying the same emotional body language and then identify the corresponding emotion being displayed in each of the videos. The results were then analyzed to determine which emotions were recognized in both cases, and how the recognition results compared.

For the videos, the actor was instructed to perform the body language descriptors in Table 7 from Chapter 5, while keeping a neutral facial expression. The actor rehearsed the body movements and postures under the guidance of the authors prior to their videotaping. With respect to the robot, the neutral pose of the robot's face was displayed throughout the videos by not actively controlling the robot's facial actuators during the display of the body language.

10.1 Participants

A total of 50 (30 female and 20 male) participants took part in the overall study after accounting for dropouts. The participants ranged in age from 17 to 63 years with a mean age of 27.78 (σ = 9.13). The participants were all from North America, where the human actor was also from. None of the participants were familiar with social robots.

10.2 Methods

Each participant logged on to a secure website that was developed by the researchers. On the website, the participants were able to watch separate videos of first the robot and then the human actor displaying the emotional body language in a random order. An initial pilot study with two groups of 10 participants was performed prior to the experiment to determine if the order of presentation of the robot and actor videos would influence recognition of the emotional body language displays. The results of a two-tailed Mann-Whitney U test performed on the recognition rates of the two groups indicated no significant order effects, U = 42, p = 0.579. Based on this finding, the videos of the robot were shown first, keeping the initial focus was on obtaining the results needed to address the first objective of the experiment, i.e. to determine if non-expert individuals would be able to identify emotions from the body language displayed by the robot. Emotional body movements and postures were displayed in the videos without any facial expressions for both the robot and actor. This procedure follows a similar approach used in other robot emotional body language studies e.g. [107, 108, 114, 117]. The robot's/actor's face was not covered when presenting the videos to the participants in order to be able to clearly show head movements and the different angles of the head that are significant descriptors for the emotions, as well as any interactions between the other body parts and the head. The participants were informed that the faces in the videos would be in a neutral emotional state. A forced-choice approach was utilized, where after the participants watched each video, they were asked to select the emotion they thought was best described in the video from the following list of eight possible emotions: sadness, fear, elated joy, surprise, anger, boredom, interest and happiness. The use of this type of forced-choice approach is very popular in studies on emotion recognition, e.g. [38, 41, 45, 114, 189]. Additionally, the forced-choice approach used herein has many advantages, including: 1) it allows for simple interpretation, i.e. it does not require the expert coding of open ended questions [202], 2) it fits the categorical nature of emotions [202], and 3) by not including a "none of the above" option, controls for participant bias, ensuring that data is collected from every participant [203, 204]. An emotion needed to be selected by the participant for each video in order for the next video to be displayed to him/her. Eight videos were each shown for the robot and for the actor.

The average length of the videos was approximately 10 seconds, during which the appropriate body movements and postures were repeated three times. Example frames from each of the videos are shown for Brian 2.1 in Figure 13 in Chapter 5 and for the human actor in Figure 31 below. The videos were recorded with a Nikon D7000 camera at 30 frames per second and a resolution of 1280 by 720 pixels. The layout of the website was such that after each video was played, the list of possible emotions was presented to the participants directly to the left of the video, as shown in Figure 32.

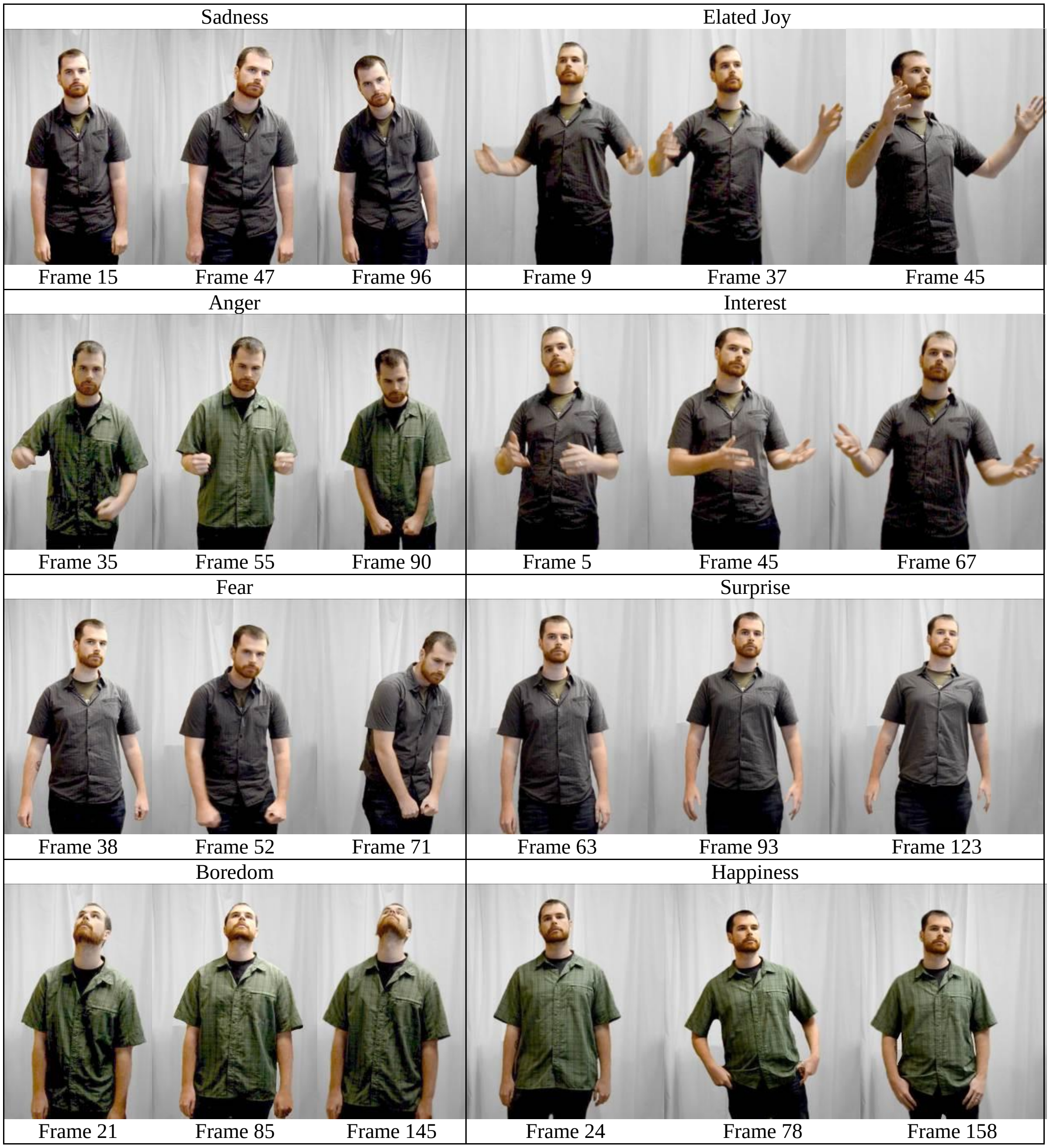


Figure 31 Example Frames of Emotional Body Language Displayed by the Human Actor for the Eight Emotions Showing Similar Movement Profiles as the Robot. [169]

Figure 32 Example of the Website Layout for the Emotional Body Language Study. [169]

10.3 Data Analysis

A within-subjects experimental design was implemented. Confusion matrices were utilized to represent the recognition rates for the emotions for both the robot and human actor. A Chi-square goodness of fit test was used to estimate the likelihood that the correct emotions that were observed for the corresponding body language did not occur due to random chance. A binomial test was utilized to determine if the desired emotion could be recognized more often than all other emotions for the respective body language.

A direct comparison study with respect to the recognition rates for the robot and human actor was conducted to determine the feasibility of using the chosen body language for the human-like social robot. A McNemar test was implemented to test if there is a significant difference between the recognition rates for the robot and the human actor. The null hypothesis used for the McNemar test was defined as: the emotion recognition rates for both the robot and human actor are the same.

10.4 Experimental Results

10.4.1 Identifying the Emotional Body Language Displayed by the Human-like Robot Brian 2.1

The recognition rates for the emotions displayed by the robot are presented in the confusion matrix in Table 28. Rows in Table 28 represent the emotions chosen by the participants and columns represent the true labeled emotions. Sadness had the highest recognition rate at 84% followed by surprise with a recognition rate of 82%. Anger and elated joy had recognition rates of 76% and 72%, while boredom and interest had rates of 56% and 38%, respectively. The emotions with the lowest recognition rates were fear with a rate of 26% and happiness with a rate of 20%. It is interesting to note that the body language for happiness was most often recognized as interest and the body language displayed for fear was recognized equally as both the emotions of fear and boredom. Interest had the highest frequency of incorrect recognitions at 11% with respect to the true labeled emotion.

Table 28 Confusion Matrix for the Emotions of the Robot.

Emotions Chosen by Participants	**Desired Emotions**								
	Sadness	Elated joy	Anger	Interest	Fear	Surprise	Boredom	Happiness	**Total**
Sadness	***42***	0	0	4	10	0	1	3	60
Elated joy	0	***36***	4	1	0	0	0	2	43
Anger	0	1	***38***	1	6	0	0	2	48
Interest	1	1	2	***19***	6	0	19	15	63
Fear	2	0	3	5	***13***	7	1	3	34
Surprise	0	1	1	1	2	***41***	0	1	47
Boredom	5	2	0	3	13	2	***28***	14	67
Happiness	0	9	2	16	0	0	1	***10***	38
Total	50	50	50	50	50	50	50	50	400

bold italic values along the diagonal refer to the number of participants that selected the desired emotion.

A chi-square goodness of fit test with $\alpha = 0.05$ was utilized to determine if the emotions recognized from the observed body language were due to random chance. The results of the chi-square test are as follows for each of the emotions:

sadness: $\chi^2(\text{df} = 7, \text{N} = 50) = 237.04$, $p < 0.001$;

elated joy: $\chi^2(\text{df} = 7, \text{N} = 50) = 171.44$, $p < 0.001$;

anger: $\chi^2(\text{df} = 7, \text{N} = 50) = 186.48$, $p < 0.001$;

interest: $\chi^2(\text{df} = 7, \text{N} = 50) = 57.20$, $p < 0.001$;

fear: $\chi^2(\text{df} = 7, \text{N} = 50) = 32.24$, $p < 0.001$;

surprise: $\chi^2(\text{df} = 7, \text{N} = 50) = 227.44$, $p < 0.001$;

boredom: $\chi^2(\text{df} = 7, \text{N} = 50) = 133.68$, $p < 0.001$; and happiness: χ^2(df = 7, N = 50) = 37.68, p <0.001.

Hence, the emotions for each of the eight displays of body language were chosen significantly above random chance.

It was hypothesized that the emotional body language movements and postures displayed by the robot would be recognized as their corresponding desired emotion more often than the other seven emotions. A binomial test was utilized to exam this hypothesis. Namely, the null hypothesis is that the desired emotion will be recognized at the same or a lower frequency than the other emotions, i.e., $p_1 \leq 0.5$. The results of the binomial test are presented in Table 29. It can be concluded that with 95% confidence the desired emotions of sadness, elated joy, anger and surprise are recognized significantly more often than any of the other emotions. A 75% confidence level was found for the emotion of boredom being recognized significantly more often than the other emotions. However, the emotions of interest, fear and happiness were not recognized significantly more often than the other emotions. Interest was the emotion most often chosen by the participants for the body language corresponding to the desired emotion of happiness.

Table 29 Results of Binomial Test for the Recognized Emotions of the Robot.

Desired Emotion	**Number of Participants that Recognized Body Language as Desired Emotion**	**Number of Participants that Recognized Body Language as Another Emotion**	**Sig** (1-tailed)	**H_0: $p_1 \leq 0.5$** α=0.05, α=0.25*
Sadness	42	8	0.000	Reject
Elated joy	36	14	0.001	Reject
Anger	38	12	0.000	Reject
Interest	19	31	0.968	Accept
Fear	13	37	> 0.999	Accept
Surprise	41	9	0.000	Reject
Boredom	28	22	0.240	Reject*
Happiness	10	40	> 0.999	Accept

10.4.2 Identifying the Emotional Body Language Displayed by the Human Actor

The recognition results for the emotions displayed by the human actor are presented in the confusion matrix in Table 30. As can be seen by the results, anger had the highest recognition rate of 100% followed by boredom which had a recognition rate of 86%. Emotions such as fear, surprise, elated joy and interest had recognition rates of 70%, 66%, 60%, and 56%, respectively. The emotions with the lowest recognition rates were sadness with a rate of 34% and happiness with a rate of only 2%. Similar to the recognition rates with respect to the robot, happiness was again considered to be the least recognized emotion from the corresponding body language. Only one participant chose happiness based on its described body language. From the results, it can be seen that the body language for sadness and happiness were more often recognized as boredom. Hence, boredom had the highest frequency of incorrect recognitions across all the emotions at 18.5%.

Table 30 Confusion Matrix for the Emotions of the Human Actor.

Emotions Chosen by Participants	**Desired Emotions**								
	Sadness	Elated joy	Anger	Interest	Fear	Surprise	Boredom	Happiness	**Total**
Sadness	***17***	0	0	2	12	0	0	5	36
Elated joy	0	***30***	0	8	0	0	0	0	38
Anger	0	1	***50***	0	0	2	1	0	54
Interest	1	9	0	***28***	0	0	6	5	49
Fear	0	0	0	0	***35***	15	0	0	50
Surprise	1	1	0	0	2	***33***	0	0	37
Boredom	31	1	0	2	1	0	***43***	39	117
Happiness	0	8	0	10	0	0	0	***1***	19
Total	50	50	50	50	50	50	50	50	400

bold italic values along the diagonal refer to the number of participants that selected the desired emotion.

A chi-square goodness of fit test with $\alpha = 0.05$ was implemented to determine if the observed emotions were chosen at a rate higher than random chance for the human actor. The test was applied to all the emotions except anger, as anger had a 100% recognition rate for the human actor. The results of the chi-square test are as follows:

sadness: $\chi^2(df = 7, N = 50) = 150.32$, $p < 0.001$;

elated joy: $\chi^2(df = 7, N = 50) = 117.68$, $p < 0.001$;

interest: $\chi^2(df = 7, N = 50) = 102.96$, $p < 0.001$;

fear: $\chi^2(df = 7, N = 50) = 169.84$, $p < 0.001$;

surprise: $\chi^2(df = 7 , N = 50) = 160.88$, $p < 0.001$;

boredom: $\chi^2(df = 7, N = 50) = 251.76$, $p < 0.001$; and

happiness: $\chi^2(df = 7, N = 50) = 201.52$, $p < 0.001$.

Hence, the emotions for these seven displays of body language were chosen significantly above random chance.

Similar to the robot emotions, it was hypothesized that the emotional body language features displayed by the actor would be recognized as their corresponding desired emotion more often than the other seven emotions. The results of the binomial test are presented in Table 31. With 95% confidence the desired emotions of anger, fear, surprise and boredom can be recognized significantly more often than any of the other emotions. Confidence levels of 89% and 75% were found for the desired emotions of elated joy and interest, respectively. On the other hand, the emotions of sadness and happiness were not recognized significantly more often than the other emotions. In particular, for both the desired emotions of happiness and sadness, the most recognized emotion by the participants, based on the corresponding body language, was boredom.

Table 31 Results of Binomial Test for the Recognized Emotions of the Human Actor.

Desired Emotion	Number of Participants that Recognized Body Language as Desired Emotion	Number of Participants that Recognized Body Language as Another Emotion	Sig (1-tailed)	H_0: $p_1 \leq 0.5$ α=0.05, α=0.11$^+$, α=0.25*
Sadness	17	31	0.968	Accept
Elated joy	30	20	0.101	Reject$^+$
Anger	50	0	0.000	Reject
Interest	28	22	0.240	Reject*
Fear	35	15	0.003	Reject
Surprise	33	17	0.016	Reject
Boredom	43	7	0.000	Reject
Happiness	1	49	> 0.999	Accept

10.4.3 Comparison

Figure 33 presents a direct comparison for the emotion recognition rates for the robot and human actor. From the figure, it can be seen that the recognition rates were higher for the human actor for the emotions of anger, interest, fear and boredom, while the robot had higher recognition rates for the emotions of sadness, elated joy, surprise and happiness.

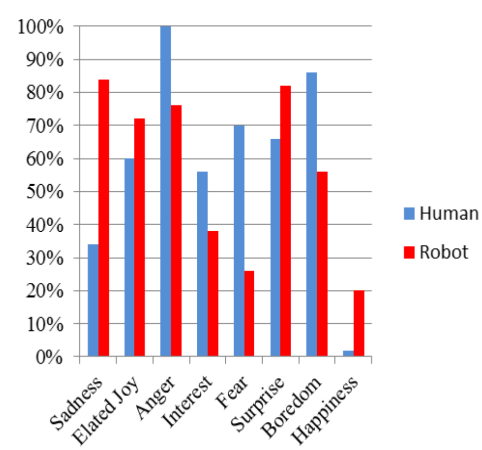


Figure 33 Comparison of Recognition Rates for Robot and Human Actor. [169]

McNemar's two-tailed test for paired proportions was used to statistically compare the recognition results from the robot and human actor. The null hypothesis was defined as the difference between the recognition rates, p_1 for the human actor and p_2 for the robot, should be zero. The first alternative hypothesis was defined as the emotion recognition rates of the body language for the human actor are higher than for the robot, and the second alternative hypothesis was defined as the recognition rates for the robot are higher than for the human actor. The 2 x 2 contingency tables comparing the recognition results of the desired emotions of the robot and actor with respect to the other emotions are presented in Table 32 with the McNemar test results presented in Table 33. Significance testing was conducted using α = 0.05. The emotions for which the null hypothesis was accepted were elated joy, interest and surprise. Hence, there was no statistical difference between the recognition rates for the robot and human actor for these emotions. For all other emotions, the null hypothesis was rejected. In particular, statistically, there is a significant difference in the recognition results for the robot and human actor for the five remaining emotions. Namely, the robot has higher recognition rates for sadness and happiness, while the human actor has higher recognition rates for the emotions of anger, fear and boredom.

Table 32 Contingency Tables for the Recognition Results of both the Robot and Human. Actor

Sadness		**Robot**		
		Sadness	Other Emotions	Total
Human	Sadness	13	4	17
	Other Emotions	29	4	33
	Total	42	8	50

Elated Joy		**Robot**		
		Elated Joy	Other Emotions	Total
Human	Elated Joy	23	7	30
	Other Emotions	13	7	20
	Total	36	14	50

Anger		**Robot**		
		Anger	Other Emotions	Total
Human	Anger	38	12	50
	Other Emotions	0	0	0
	Total	38	12	50

Interest		**Robot**		
		Interest	Other Emotions	Total
Human	Interest	13	15	28
	Other Emotions	6	16	22
	Total	19	31	50

Fear		**Robot**		
		Fear	Other Emotions	Total
Human	Fear	13	22	35
	Other Emotions	0	15	15
	Total	13	37	50

Surprise		**Robot**		
		Surprise	Other Emotions	Total
Human	Surprise	29	4	33
	Other Emotions	12	5	17
	Total	41	9	50

Boredom		**Robot**		
		Boredom	Other Emotions	Total
Human	Boredom	28	15	43
	Other Emotions	0	7	7
	Total	28	22	50

Happiness		**Robot**		
		Happiness	Other Emotions	Total
Human	Happiness	0	1	1
	Other Emotions	10	39	49
	Total	10	40	50

Table 33 McNemar Significance Results for the Robot and Human Actor Recognition Rates.

Emotion	**McNemar Sig** (2-tailed)	**H_0: $p_1 - p_2 = 0$** (α=0.05)	**Subject Most Recognized**
Sadness	0.000	Reject	Robot
Elated joy	0.263	Accept	Same
Anger	0.000	Reject	Human
Interest	0.078	Accept	Same
Fear	0.000	Reject	Human
Surprise	0.077	Accept	Same
Boredom	0.000	Reject	Human
Happiness	0.012	Reject	Robot

10.5 Discussions

The recognition results for the human-like social robot showed that participants were able to recognize the emotional body language for sadness, elated joy, anger, surprise and boredom, as defined by Wallbott [47] and de Meijer [48], with rates over 55%. All these emotions had recognition rates significantly above random chance with respect to all other emotions for the same body language. The body language for the emotion of fear was recognized by the participants both as fear and boredom with the exact same frequency. This can be a difficult emotion for the robot to express based on the defined body movements and postures due to the rigidity of the robot's body. For example, the rigid body of the robot does not allow it to easily curl in the shoulders and bend the back similar to how a human would for this particular emotion, i.e., Figure 31. Furthermore, it is difficult for the robot to mimic the tensing of the muscles in the body to represent the force and energy of the high dynamic movements for this particular emotion. This made the recognition of this emotion more challenging for the participants. Furthermore, as the emotional body language for fear required the robot to turn its head away and bow its trunk, some participants confused this as the robot was displaying boredom.

For the robot, the body language for the desired emotion of happiness was recognized more often as interest and boredom. For the actor, the body language for the desired emotion of happiness was recognized most often as boredom. These other emotions contain similar descriptors to happiness such as stretching the trunk and low movement dynamics which could contribute to the confusion. In general, the body language for happiness had low recognition rates for both the robot and human actor, although the recognition rates were significantly higher for the robot than the actor. Unlike the robot, during this body language display, the actor also had his hands in his pockets for approximately half of the duration of the video. Hands in the pockets have been found to be perceived as a number of different affective states including calm and easygoing [205], casual attitude [206], relief [207], and sad [207]. Hence, this particular gesture may have also resulted in the majority of the participants recognizing this body language display as boredom for the actor. The similarity in descriptors can also be the reason why the robot's emotional body language for interest was recognized as happiness by 32% of the participants. Hence, alternative body language descriptors may need to be considered and tested for the emotion of happiness. The challenge will be to identify potential descriptors for happiness for the robot that will also be unique from those used for elated joy, where both emotions have positive valence, but the latter has higher arousal. Wallbott [47] is the only researcher that provides specific human body language descriptors for the emotions of happiness, elated joy, boredom and interest. In this study, Wallbott's descriptors were used for the first three emotions and descriptors from de Meijer for the emotion of interest. For interest, Wallbott's body language descriptors are similar to those defined by de Meijer, with the exception that de Meijer also included descriptors that describe the direction and dynamics of body movements for this emotion. The inclusion of other modes such as facial expressions may also need to be considered for happiness. For example, it has been shown in several studies that a universal human facial expression for happiness includes such descriptors as raising the checks and moving the corners of the mouth upwards [208], hence adding such descriptors to the body language for happiness might be necessary in order to increase recognition rates for this particular emotion for the robot.

The recognition results for the actor showed that the participants most often associated the body language for sadness with boredom, however, this was not the case for the robot. For the robot, the body language for the desired emotion of sadness was recognized significantly more often as sadness than any of the other emotions. From the comparison study, it was determined that the desired emotion of sadness was recognized at significantly higher rates for the robot than the actor. This may be a result of the difference in the head positions of the robot versus the actor during the videos. On average, the robot's head was facing more downwards than the actor's head while displaying the body movements for sadness, as the robot was not able to slouch its shoulders. Studies by both Darwin [209] and Bull [210] have found that dropping/hanging the head is related to the emotion of sadness. The emotions of interest and surprise had statistically similar recognition results for the robot and actor; this was due to the fact that the robot was able to easily replicate the body movements for these emotions and did so in a similar manner as the actor did. For the emotion of elated joy, due to each shoulder of the robot having one fewer rotational degree of freedom than a human, the robot generated the opening and upwards arm movements by also moving its upper arms outwards compared to the actor who directly lifted the upper arms forwards. Despite this difference, the recognition results were statistically similar to that of the human actor. The emotional body language for anger, boredom and fear were recognized at statistically higher recognition rates for the actor, this can be a result of the robot not being able to directly mimic the tensing of the muscles (for angry and fear) or curling in the shoulders and bending the back (for boredom and fear) as previously mentioned.

As both the robot's and actor's faces were visible in the videos, the lack of facial expressions could have influenced the recognition rates for the emotions, even though the participants were informed that only emotional body language without any facial expressions was displayed in both sets of videos. Namely, this might have been a reason why happiness had low recognition rates for both the robot and actor. This could have also been the cause of the confusion for the emotion of fear being recognized as both fear and boredom when the corresponding body language was displayed by the robot. Since the robot's eyes did not move independently of the head, the robot did not keep eye contact with the camera to the same extent as the actor did for the emotional body language displays of sadness and surprise. For the display of sadness, due to its more downwards head pose, the robot averted its gaze from the camera for 89% of the video, while the actor averted his gaze from the camera for 55% of the video. As previously mentioned, this more downwards head pose of the robot and therefore its averted gaze may be a result of why its display of sadness had a higher recognition rate. For the display of surprise, due to the range of motion of the robot's body, the robot averted its gaze for 95% of the video, while the actor did not avert his eyes. Despite this difference in eye gaze, the recognition rate for the robot for surprise was statistically similar to the recognition rate for the actor. Although when comparing Figure 13 in Chapter 5 and Figure 31, the robot's body language for fear and happiness appear to have slightly more instances of averted gaze in comparison to the human, the overall amount of time that Brian 2.1 and the actor had averted gazes during their respective videos for these emotions was within 10% of each other. Previous studies have shown that eye gaze direction does not directly influence recognition of emotions displayed by facial expressions [211] and that they are processed independently [212], however, to the authors' knowledge, there have been no studies that have investigated the direct influence of eye gaze on the recognition of emotional body language. Therefore, this relationship should be further explored in future work.

The recognition rates for the robot were also compared to the recognition rates that Wallbott obtained in [47] for the same body language descriptors used for happiness, sadness, boredom and elated joy to provide further insight. Unfortunately, a similar comparison could not be conducted with the emotions obtained from de Meijer's descriptors as recognition rates were not provided in [48]. The recognition rates of the emotions elated joy and boredom of Brian 2.1 were found to be within 10% of the recognition results that Wallbott observed for these emotions in [47]. Sadness also had high recognition rates for this study and Wallbott's study. In [47], happiness had a good recognition rate, being distinguishable from all the other emotions except for contempt, which is an emotion that was not considered for this study.

Overall, the experimental results showed that the body language descriptors were effective in displaying the emotions of sadness, elated joy, anger, surprise and boredom for the social human-like robot Brian 2.1, warranting the potential use of these social emotions and corresponding body language for the robot in natural and social HRI settings. On the other hand, the body language for the emotions of happiness, fear and interest were not well recognized for the robot.

While previous studies have compared human and artificial displays of emotional facial expressions and have shown that the later can also be recognized effectively (though with lower recognition rates than the human) [213-215], the comparison study is novel in that it focuses on a robot's display of emotional body language. In general, the work presented in this chapter can be used as a reference when determining the emotional body language of other life-sized human-like robots or androids. With respect to android body language, it has been stated that there has been little active research in this area [216].

10.6 Summary

This chapter investigated robotic affective communication as displayed through body language. Namely, it investigated the use of emotional body language for the human-like social robot Brian 2.1 utilizing body movement and posture descriptors identified in human emotion research. Experiments were conducted to determine: 1) if non-expert individuals would be able to identify the eight social emotions of sadness, fear, elated joy, surprise, anger, boredom, interest and happiness from the display of Brian 2.1's body language which has been derived from a combination of human body language descriptors, and 2) compare how individuals interpret the same emotional body language descriptors displayed by the social robot with fewer degrees of freedom and a human actor, in order to determine if the desired emotions can be communicated by the robot as effectively as by a human. Experimental results showed that participants were able to recognize the robot's emotional body language for sadness, elated joy, anger, surprise and boredom with high recognition rates. Even though the robot was not able to implement some body movement features due to its rigid body, the participants were still able to recognize the majority of the emotions. When comparing the recognition rates, it was determined that the emotion of sadness was even recognized at significantly higher rates for the robot than the human actor, while the robot and actor had similar recognition rates for elated joy, surprise and interest. Both the robot and actor had the lowest recognition rates for the emotion of happiness, due to its similarity in body movement features to other emotions. Only the emotions of anger, fear and boredom were recognized at a significantly higher rate for the human actor. Overall, these experimental findings demonstrate that certain human-based body movements and postures that can represent social emotions can be effectively displayed by a life-sized human-like robot.

Chapter 11
Conclusions and Recommendations for Future Work

This chapter presents a summary of the research challenges addressed in this thesis as well as the systems and HRI experiments designed and developed to address them. The work encompasses unique methods to autonomously detect affect from human body language during HRI and design affective body language for human-like socially assistive robots. Additionally, possible directions for future research are also discussed in this chapter.

11.1 Summary of Contributions

This thesis has: i) developed the first compact real-time multimodal perception systems to identify a person's affect from his/her display of natural body language during socially assistive HRI, and ii) developed affective body language for a human-like socially assistive robot allowing it to communicate emotions during social HRI. These systems will allow socially assistive robots to respond intelligently to a person's affective state during socially assistive HRI in such applications as social and cognitive therapy/interventions in order to promote engagement and complete interaction task goals.

Body language is a very important part of natural human communication and has been shown to be used to communicate a person's affective state during social interaction. The automated identification and classification of affective body language is very challenging due to the large number of configurations in a high dimensionality search space. This task is especially difficult when it is intended for a socially assistive robot that engages in one-on-one HRI in a variety of environments, necessitating the use of sensory systems mounted on the robot. In addition, developing affective body language for human-like socially assistive robots is important to allow the robot to effectively and naturally communicate its goals to a user in order to complete the interaction objectives.

The contributions of this thesis are outlined herein by reviewing the novel systems developed and experiments performed to address the aforementioned research challenges in the following subsections.

11.1.1 Detecting Affect from Static Body Language

An automated multi-modal static body language recognition and classification system was developed to determine a person's affect in terms of their accessibility (i.e., openness and rapport) towards a robot during natural one-on-one interactions - *the first of its kind*. The proposed system utilizes 3D sensory information of a person and skin region information of the lower arms and head, to generate an ellipsoid model of a person's static body language. The orientations of the ellipsoid model's upper and lower trunks as well as the pose of the arms are utilized to identify the person's accessibility towards a robot during social HRI. The proposed approach is not system dependent and has been implemented in this work with two different sets of sensors: i) a thermal camera and TOF camera; and ii) a 2D color camera and 3D depth imager.

11.1.2 Detecting Affect from Dynamic Body Language

The first automated dynamic body language recognition and classification system for a socially assistive robot to identify the affective state of seated individuals was developed. The proposed system identifies a person's affect in terms of his/her level of valence (pleasantness or unpleasantness) and arousal (level of activity) based the 3D data from a depth imager mounted on a robot. Valence and arousal levels are classified from dynamic body language features utilizing learning-based classifiers.

11.1.3 Designing Affective Body Language for a Socially Assistive Human-Like Robot

This work considered the first application of human affective body language, as identified in psychology and behavioural research, to a human-like socially assistive robot. These body language displays were mapped onto a robot to identify corresponding affective states that can be utilized by the robot during social interactions. The body language utilizes movements and postures of the trunk, head and arms as well as overall movement quality in order for a robot to display affect.

11.1.4 Socially Assistive Human-Robot Interaction Experiments

As one of the aims of this thesis is on bi-directional emotional communication between a robot and users from vulnerable populations, this thesis presents novel user studies to investigate how elderly users interact with and perceive a socially assistive robot during assistive activities.

During both cognitively stimulating memory card game interactions and the meal-eating activity, the socially assistive robot utilized prompts to help elderly users through the steps of the activities as well as used positive statements to encourage engagement in the HRI. The results of the experiments showed that the elderly users complied with the robot's prompts, engaged in the interactions, and overall enjoyed interacting with the robot. These experiments have set the framework for conducting assistive user studies with elderly adults.

11.1.5 Detecting Affect from Static Body Language Experiments

Two sets of novel experiments were performed to investigate the performance of the automated static body language recognition and classification system at identifying a person's accessibility during HRI. The first set of experiments determined that the proposed system correctly identified the accessibility levels of user's natural body language displayed while engaged in social HRI with two different sensory systems. The second set of experiments showed that the proposed system more accurately identified the accessibility of people engaged in one-on-one HRI than a popular commercially available system.

11.1.6 Detecting Affect from Dynamic Body Language Experiments

The performance of the automated dynamic body language recognition and classification system was tested with unique socially assistive HRI experiments with seated elderly participants. The results found that a variety of learning-based classifiers are capable of correctly classifying the valence and arousal levels of elderly users engaged in the meal-eating activity with a human-like socially assistive robot. The main contribution of these experiments is to show that affective human body language of these individuals can be recognized and classified by a robot.

11.1.7 Affect-Aware Robot Behaviour Experiments

Once a person's affect can be determined, this work further investigates how users interact with and perceive a robot that autonomously utilizes a user's accessibility levels throughout one-on-one socially assistive HRI to determine its own behaviours. A within subjects HRI comparison study was performed between: i) an accessibility-aware robot behaviour type that responds emotionally to a user's accessibility level and the activity state, and ii) a non-accessibility-aware robot behaviour type that only responds to the current state of the activity. The results found that the participants were more accessible towards the accessibility-aware robot behaviour type and perceived such a robot to be more socially intelligent.

11.1.8 Recognition of Affective Body Language Displayed by a Human-Like Socially Assistive Robot

Finally, novel experiments to investigate the recognition of affective body language of a human-like socially assistive robot for which the body language was designed using human psychology and behavioural research were performed. These experiments had two main goals: 1) to determine if non-experts can recognize the affective states displayed by the robot's body language, and 2) compare how these individuals interpret the same affective body language displayed by a human actor and the robot. The results indicate that some human affective body language displays can be effectively mapped onto and displayed by a human-like socially assistive robot with fewer degrees of freedom than a human.

11.2 Recommendations for Future Work

Future research directions should include investigating the detection of additional affective states through body language. This thesis focused on identifying a person's accessibility (categorical affect), and valence and arousal (dimensional affect) from body language, however additional categorical affective states such as basic emotions may be detectable from body language displays. An increased set of detectable affective states will allow socially assistive robots to respond appropriately to a larger variety of user actions.

Future research should also include developing automated affective body language recognition and classification systems, both static and dynamic, that allow for more than one person to be engaged in an interaction with a robot. This will also allow for experimental investigations into how people interact with and perceive socially assistive robots when there is a group of people participating in the interaction.

Future research could also focus on fusing affect information from a combination of multimodal means such as body language, facial expressions, tone of voice and speech, as well as non-contact physiological sensing. Sensing a number of available modes will provide a socially assistive robot with more information in order to accurately estimate a user's current affective state.

Future socially assistive HRI experiments should also focus on long-term acceptance and use to determine if elderly users maintain their engagement, compliance and enjoyment over a large number of interactions and over longer periods of time. This will help further validate the use of such robots for long-term use as assistants for the elderly in nursing homes and long-term care facilities.

Appendix A
List of My Publications

Book Chapters

[1] D. McColl and G. Nejat, “Human-Robot Interaction for Assistance with Activities of Daily Living: A Case Study of the Socially and Cognitively Engaging Brian 2.1 in the Long-Term Care Setting,” *Speech and Automata in Healthcare: Voice Controlled Medical and Surgical Robots,* De Gruyter Publishing, in Print.

[2] D. McColl and G. Nejat, “A Human Affect Recognition System for Socially Interactive Robots,” *In the Handbook of Research on Technoself: Identity in a Technological Society,* IGI Global Publishing, pp. 554-573, 2013.

Journal Publications

[3] W. G. Louie, D. McColl and G. Nejat, “Acceptance and Attitudes Towards a Human-Like Socially Assistive Robot by Older Adults,” *Assistive Technology*, In Print.

[4] D. McColl and G. Nejat, “A Socially Assistive Robot that can Monitor Affect of the Elderly during Meal-Time Assistance,” *Journal of Medical Devices-Transactions of the ASME*, 2014, DOI:10.1115/1.4027109.

[5] D. McColl and G. Nejat, “Recognizing Emotional Body Language Displayed by a Human-like Social Robot”, *International Journal of Social Robotics*, Vol. 6, No. 2, pp. 261-280, 2014.

[6] D. McColl and G. Nejat, “A Human-Robot Interaction Study with a Meal-Time Socially Assistive Robot and Older Adults at a Long-term Care Facility,” *Special Issue on HRI System Studies in the Journal of Human-Robot Interaction*, Vol. 2, No. 1, pp. 152-171, 2013.

[7] D. McColl, W. G. Louie, G. Nejat, “Brian 2.1: A Socially Assistive Robot for the Elderly and Cognitively Impaired,” *Special Issue on Assistive Robotics, IEEE Robotics & Automation Magazine*, Vol. 20, No. 1, March 2013.

[8] D. McColl and G. Nejat, “A Socially Assistive Robot that can Interpret Affective Body Language During One-on-One Human-Robot Interactions,” *International Journal of Biomechatronics and Biomedical Robotics*, Vol. 2, No. 1, pp. 39-49, 2012.

[9] D. McColl, J. Chan and G. Nejat, "A Socially Assistive Robot for Meal-Time Cognitive Interventions," *Journal of Medical Devices-Transactions of the ASME,* Vol. 6, No. 1, 017559, 2012.

[10] D. McColl, Z. Zhang, and G. Nejat, "Human Body Pose Interpretation and Classification for Human-Robot Interaction," *International Journal of Social Robotics*, Vol. 3, No. 3, pp. 313-332, 2011.

Conference Proceedings

[11] D. McColl and G. Nejat “Determining the Affective Body Language of Older Adults during Socially Assistive HRI,” *IEEE International Conference on Intelligent Robots and Systems (IROS)*, 2014, Accepted.

[12] 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," *IEEE International Symposium on Robot and Human Interactive Communication*, pp. 345-350, Paris, France, Sept. 2012.

[13] D. McColl and G. Nejat, "Affect Detection from Body Language during Social HRI," *IEEE International Symposium on Robot and Human Interactive Communication*, pp. 1013-1018, Paris, France, Sept. 2012.

[14] D. McColl and G. Nejat, "A Socially Assistive Robot That Can Interpret Human Body Language," *ASME International Design Engineering Technical Conferences & Computers and Information in Engineering Conference*, Washington, DC, DETC2011-48031, 2011.

References

[1] T. K. Tatemichi et al., "Cognitive impairment after stroke: frequency, patterns, and relationship to functional abilities," *J. Neurol. Neurosurg. Psychiatry.*, vol. 57, pp. 202-207, Feb, 1994.

[2] M. Prince, M. Guerchet and M. Prina, "Policy brief for heads of government: The global impact of dementia 2013–2050," *London: Alzheimers Disease Int.*, 2013.

[3] E. F. LoPresti, R. C. Simpson, N. Kirsch, D. Schreckenghost and S. Hayashi, "Distributed cognitive aid with scheduling and interactive task guidance," *J. Rehabilitation Res. and Develop.*, vol. 45, pp. 505-521, 2008.

[4] G. W. Small, et al., "Diagnosis and treatment of Alzheimer disease and related disorders: consensus statement of the American Association for Geriatric Psychiatry, the Alzheimer's Association, and the American Geriatrics Society," *J. Am. Med. Assoc.*, vol. 278, pp. 1363-1371, 1997.

[5] M. J. Mataric, A. Okamura and H. Christensen, "A Research Roadmap for Medical and Healthcare Robotics," *NSF/CCC/CRA Roadmapping for Robotics Workshop*, Arlington/Washington, DC, 2008, pp. 1-30.

[6] H. X. Wang, A. Karp, B. Winblad and L. Fratiglioni, "Late-life engagement in social and leisure activities is associated with a decreased risk of dementia: a longitudinal study from the Kungsholmen project," *Am. J. Epidemiol.*, vol. 155, pp. 1081-1087, Jun 15, 2002.

[7] M. Eisele et al., "Influence of social support on cognitive change and mortality in old age: results from the prospective multicentre cohort study AgeCoDe," *BMC Geriatr.*, vol. 12, pp. 9-2318-12-9, Mar 20, 2012.

[8] K. Ball, et al., "Effects of cognitive training interventions with older adults: a randomized controlled trial," *J. Am. Med. Assoc.*, vol. 288, pp. 2271-2281, 2002.

[9] D. Sitzer, E. Twamley and D. Jeste, "Cognitive training in Alzheimer's disease: a meta-analysis of the literature," *Acta Psychiatr. Scand.*, vol. 114, pp. 75-90, 2006.

[10] M. J. Weiss, K. Fiske and S. Ferraioli, "Treatment of autism spectrum disorders," in *Treating Childhood Psychopathology and Developmental Disabilities*, LNCS, Springer, 2009, pp. 287-332.

[11] E. Aguirre et al., "Maintenance Cognitive Stimulation Therapy (CST) for dementia: a single-blind, multi-centre, randomized controlled trial of Maintenance CST vs. CST for dementia," *Trials*, vol. 11, pp. 46-6215-11-46, Apr 28, 2010.

[12] T. Hashimoto and H. Kobayashi, "Study on natural head motion in waiting state with receptionist robot SAYA that has human-like appearance," in *IEEE Workshop on Robotic Intell. in Informationally Structured Space*, 2009, pp. 93-98.

[13] A. Haasch et al., "BIRON - the bielefeld robot companion," in *IEEE Int. Workshop on Advances in Service Robots*, 2004, pp. 1-6.

[14] F. Belkhouche, B. Belkhouche and P. Rastgoufard, "Line of sight robot navigation toward a moving goal," *IEEE Trans. Syst., Man Cybern. B, Cybern.*, vol. 36, pp. 255-267, 2006.

[15] D. Feil-Seifer, K. Skinner and M. J. Matarić, "Benchmarks for evaluating socially assistive robotics," *Interaction Stud.*, vol. 8, pp. 423-439, 2007.

[16] G. Nejat, Y. Sun and M. Nies, "Assistive robots in health care settings," *Home Health Care Management & Practice*, vol. 21, pp. 177-187, 2009.

[17] T. Shibata and K. Wada, "Robot therapy: a new approach for mental healthcare of the elderly - a mini-review," *Gerontology*, vol. 57, pp. 378-386, 2011.

[18] M. Montemerlo, J. Pineau, N. Roy, S. Thrun and V. Verma, "Experiences with a mobile robotic guide for the elderly," in *AAAI Nat. Conf. Artificial Intell.*, 2002, pp. 587-592.

[19] J. Fasola and M. J. Mataric, "Robot exercise instructor: A socially assistive robot system to monitor and encourage physical exercise for the elderly," in *IEEE Int. Symp. Robot and Human Interactive Commun.*, 2010, pp. 416-421.

[20] A. Tapus, C. Tapus and M. Matarić, "Long term learning and online robot behavior adaptation for individuals with physical and cognitive impairments," in *Field and Service Robotics*, 2010, pp. 389-398.

[21] T. Zhang et al., "Service robot feature design effects on user perceptions and emotional responses," *Intell. Service Robotics*, vol. 3, pp. 73-88, 2010.

[22] M. Heerink, B. Krose, V. Evers and B. Wielinga, "Measuring acceptance of an assistive social robot: A suggested toolkit," in *IEEE Int. Symp. Robot and Human Interactive Commun.*, 2009, pp. 528-533.

[23] T. Klamer and S. Allouch, "Zoomorphic robots used by elderly people at home," in *ACM Int. Conf. on Human Factors in Computing Syst.*, *ACM Press*, 2010, pp. 1-2.

[24] A. Tapus, M. Maja and B. Scassellatti, "The grand challenges in socially assistive robotics," *IEEE Robotics and Automation Mag.*, vol. 14, 2007, pp. 35-42.

[25] C. Breazeal, "Social Interaction in HRI: The Robot View," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 34, pp. 181-186, 2004.

[26] M. Xin and E. Sharlin, "Playing games with robots - A method for evaluating human-robot interaction," in *Human-Robot Interaction*, N. Sankar, Ed. Vienna, Austria: 2007, pp. 469-480.

[27] T. Fong, I. Nourbakhsh and K. Dautenhahn, "A survey of socially interactive robots," *Robotics and Autonomous Syst.*, vol. 42, pp. 143-166, 2003.

[28] D. Morris, P. Collett, P. Marsh and M. Shaughnessy, *Gestures, their Origin and Distribution.* London: Jonathan Cape Ltd, 1979.

[29] S. Gong, P. W. McOwan and C. Shan, "Beyond facial expressions: Learning human emotion from body gestures," in *Br. Machine Vision Conf.*, 2007, pp. 1-10.

[30] P. Ekman and V. Friesen, "Head and Body Cues in the Judgement of Emotion: A Reformation," *Perceptual and Motor Skills,* vol. 24, pp. 711-724, 1967.

[31] A. Mehrabian, "Significance of Posture and Position in the Communication of Attitude and Status Relationships," *Psychol. Bulletin,* vol. 71, pp. 359-372, 1969.

[32] P. Ekman and W. V. Friesen, "The repertoire of nonverbal behavior: Categories, origins, usage, and coding," *Nonverbal Commun., Interaction, and Gesture,* pp. 57-106, 1981.

[33] W. T. James, "A study of the expression of bodily posture," *J. General Psychol.*, vol. 7, pp. 405-437, 1932.

[34] K. L. Walters and R. D. Walk, "Perception of emotion from body posture," *Bulletin of Psychonomic Soc.*, vol. 24, pp. 329-329, 1986.

[35] S. J. Schouwstra and J. Hoogstraten, "Head Position and Spinal Position as Determinant of Perceived Emotional State," *Perceptual and Motor Skills,* vol. 81, pp. 673-674, 1995.

[36] M. Coulson, "Attributing Emotion to Static Body Postures: Recognition, Accuracy, Confusions, and Viewpoint Dependence," *J. Nonverbal Behavior,* vol. 28, pp. 117-139, 2004.

[37] A. S. Shaarani and D. M. Romano, "Perception of emotions from static postures," in *Affective Computing and Intell. Interaction,* LNCS, Springer, 2007, pp. 761-762.

[38] B. de Gelder and J. Van den Stock, "The Bodily Expressive Action Stimulus Test (BEAST). Construction and Validation of a Stimulus Basis for Measuring Perception of Whole Body Expression of Emotions," *Front. Psychol.,* vol. 2, pp. 181, Aug 9, 2011.

[39] H. K. Meeren, C. C. van Heijnsbergen and B. de Gelder, "Rapid perceptual integration of facial expression and emotional body language," *Proc. Natl. Acad. Sci. U.S.A.,* vol. 102, pp. 16518-16523, Nov 8, 2005.

[40] J. Van den Stock, R. Righart and B. De Gelder, "Body expressions influence recognition of emotions in the face and voice." *Emotion,* vol. 7, pp. 487, 2007.

[41] A. Atkinson, W. Dittrich, A. Gemmell and A. Young, "Emotion perception from dynamic and static body expression in point-light and full-light displays," *Perception,* vol. 33, pp. 717-746, 2004.

[42] J. Montepare, S. Goldstein and A. Clausen, "The identification of emotions from gait information," *J. Nonverbal Behavior,* vol. 11, pp. 33-42, 1987.

[43] S. Brownlow, A. R. Dixon, C. A. Egbert and R. Radcliffe, "Perception of movement and dancer characteristics from point-light displays of dance." *The Psychol. Rec.,* vol. 47, pp. 411-421, 1997.

[44] J. Montepare, E. Koff, D. Zaitchik and M. Albert, "The Use of Body Movements and Gestures as Cues to Emotions in Younger and Older Adults," *J. Nonverbal Behavior,* vol. 23, pp. 133-152, 1999.

[45] F. Pollick, H. Paterson, A. Bruderlin and A. Sanford, "Perceiving affect from arm movement," *Cognition,* vol. 82, pp. B51-B61, 2001.

[46] J. A. Russell, "A circumplex model of affect," *J. Personality and Soc. Psychol.,* vol. 39, pp. 1161-1178, 1980.

[47] H. G. Wallbott, "Bodily Expression of Emotion," *Eur. J. Soc. Psychol.,* vol. 28, pp. 879-896, 1998.

[48] M. de Meijer, "The Contribution of General Features of Body Movement to the Attribution of Emotions," *J. Nonverbal Behavior,* vol. 13, pp. 247-268, 1989.

[49] S. Frintrop, A. Königs, F. Hoeller and D. Schulz, "A component-based approach to visual person tracking from a mobile platform," *Int. J. Soc. Robotics,* vol. 2, pp. 53-62, 2010.

[50] Y. Benezeth, B. Emile, H. Laurent and C. Rosenberger, "Vision-based system for human detection and tracking in indoor environment," *Int. J. Soc. Robotics,* vol. 2, pp. 41-52, 2010.

[51] C. Juang, C. Chang, J. Wu and D. Lee, "Computer vision-based human body segmentation and posture estimation," *IEEE Trans. Syst., Man Cybern. A, Syst. Humans*, vol. 39, pp. 119-133, 2009.

[52] G. Mori, X. Ren, A. A. Efros and J. Malik, "Recovering human body configurations: Combining segmentation and recognition," in *IEEE Comput. Soc. Conf. Comput. Vision and Pattern Recognition*, 2004, pp. II-326-II-333 Vol. 2.

[53] B. Chen, N. Nguyen and G. Mori, "Human pose estimation with rotated geometric blur," in *IEEE Workshop on Applicat. of Comput. Vision*, 2008, pp. 1-6.

[54] Q. Pham, L. Gond, J. Begard, N. Allezard and P. Sayd, "Real-time posture analysis in a crowd using thermal imaging," in *IEEE Conf. Comput. Vision and Pattern Recognition*, 2007, pp. 1-8.

[55] M. B. Holte, T. B. Moeslund and P. Fihl, "View invariant gesture recognition using the CSEM SwissRanger SR-2 camera," *Int. J. Intell. Syst. Technol. and Applicat.*, vol. 5, pp. 295-303, 2008.

[56] D. Demirdjian and C. Varri, "Driver pose estimation with 3D time-of-flight sensor," in *IEEE Workshop on Comput. Intell. in Vehicles and Vehicular Syst.*, 2009, pp. 16-22.

[57] P. Kohli, J. Rihan, M. Bray and P. H. Torr, "Simultaneous segmentation and pose estimation of humans using dynamic graph cuts," *Int. J. Computer Vision*, vol. 79, pp. 285-298, 2008.

[58] A. Gupta, A. Mittal and L. S. Davis, "Constraint integration for efficient multiview pose estimation with self-occlusions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, pp. 493-506, 2008.

[59] M. Van den Bergh, E. Koller-Meier and L. Van Gool, "Real-time body pose recognition using 2D or 3D haarlets," *Int. J. Computer Vision*, vol. 83, pp. 72-84, 2009.

[60] S. Y. Cheng, S. Park and M. M. Trivedi, "Multiperspective thermal ir and video arrays for 3d body tracking and driver activity analysis," in *IEEE Comput. Soc. Conf. Comput. Vision and Pattern Recognition*, 2005, pp. 3-3.

[61] S. Knoop, S. Vacek and R. Dillmann, “A Human Body Model for Articulated 3D Pose Tracking,” in *Humanoid Robots, New Developments, Advanced Robotic Systems International*, Pina Filho (ed), Croatia, pp.505-520, 2007.

[62] Microsoft, Kinect for Windows programming guide, 2014, available http://msdn.microsoft.com/en-us/library/hh855348.aspx.

[63] M. Shimizu, T. Yoshizuka and H. Miyamoto, "A gesture recognition system using stereo vision and arm model fitting," in *Int. Congr. Series*, 2007, pp. 89-92.

[64] M. Hasanuzzaman, T. Zhang, V. Ampornaramveth and H. Ueno, "Gesture-based human-robot interaction using a knowledge-based software platform," *Industrial Robot: An Int. J.*, vol. 33, pp. 37-49, 2006.

[65] M. Hasanuzzaman, V. Ampornaramveth, T. Zhang, M. Bhuiyan, Y. Shirai and H. Ueno, "Real-time vision-based gesture recognition for human robot interaction," in *IEEE Int. Conf. on Robotics and Biomimetics*, 2004, pp. 413-418.

[66] H. S. Park et al., "HMM-based gesture recognition for robot control," in *Pattern Recognition and Image Analysis*, LNCS, Springer, 2005, pp. 607-614.

[67] V. Bonato et al., "A real time gesture recognition system for mobile robots." in *Int. Conf. Informatics in Control, Automation and Robotics*, 2004, pp. 207-214.

[68] S. Waldherr, R. Romero and S. Thrun, "A gesture based interface for human-robot interaction," *Autonomous Robots*, vol. 9, pp. 151-173, 2000.

[69] S. I. Bahadori, L. Nardi and G. D Settembre, "Stereo vision based human body detection from a localized mobile robot," in IEEE Conf. Advanced Video and Signal Based Surveillance, 2005, pp. 499.

[70] B. Burger, I. Ferrané and F. Lerasle, "Multimodal interaction abilities for a robot companion," in *Comput. Vision Syst.*, LNCS, Springer, 2008, pp. 549-558.

[71] F. Guan, L. Li, S. S. Ge and A. P. Loh, "Robust human detection and identification by using stereo and thermal images in human robot interaction," *Int. J. Information Acquisition*, vol. 4, pp. 161-183, 2007.

[72] N. Werghi, "Segmentation and modeling of full human body shape from 3-D scan data: A survey," *IEEE Tran. Syst., Man, Cybern. C, Appl. Rev.*, vol. 37, pp. 1122-1136, 2007.

[73] T. B. Moeslund and E. Granum, "A survey of computer vision-based human motion capture," *Comput. Vision Image Understanding*, vol. 81, pp. 231-268, 2001.

[74] T. B. Moeslund, A. Hilton and V. Krüger, "A survey of advances in vision-based human motion capture and analysis," *Comput. Vision Image Understanding*, vol. 104, pp. 90-126, 2006.

[75] I. Cohen, A. Garg and T. S. Huang, "Emotion recognition from facial expressions using multilevel HMM," in *Neural Information Processing Systs*, 2000.

[76] M. Madsen, R. El Kaliouby, M. Goodwin and R. Picard, "Technology for just-in-time in-situ learning of facial affect for persons diagnosed with an autism spectrum disorder," in *Int. ACM SIGACCESS Conf. Comput. and Accessibility,* 2008, pp. 19-26.

[77] C. J. Duthoit, T. Sztynda, S. K. Lal, B. T. Jap and J. I. Agbinya, "Optical flow image analysis of facial expressions of human emotion: Forensic applications," in *Int. Conf. on Forensic Appl. and Techniques in Telecommunications, Information, and Multimedia and Workshop,* 2008, pp. 5.

[78] M. N. Dailey, G. W. Cottrell, C. Padgett and R. Adolphs, "EMPATH: A neural network that categorizes facial expressions," *J. Cogn. Neurosci.,* vol. 14, pp. 1158-1173, 2002.

[79] C. L. Lisetti and A. Marpaung, "Affective cognitive modeling for autonomous agents based on scherer's emotion theory," in *Advances in Artificial Intell.*, LNCS, Springer, 2007, pp. 19-32.

[80] Y. Tian, T. Kanade and J. F. Cohn, "Facial expression analysis," in *Handbook of Face Recognition*, Springer, 2005, pp. 247-275.

[81] L. Kessous, G. Castellano and G. Caridakis, "Multimodal emotion recognition in speech-based interaction using facial expression, body gesture and acoustic analysis," *J. Multimodal User Interfaces,* vol. 3, pp. 33-48, 2010.

[82] H. Gunes and M. Piccardi, "Affect recognition from face and body: Early fusion vs. late fusion," in *IEEE Int. Conf. Syst., Man Cybern.,* 2005, pp. 3437-3443.

[83] K. Schindler, L. Van Gool and B. de Gelder, "Recognizing emotions expressed by body pose: A biologically inspired neural model," *Neural Networks,* vol. 21, pp. 1238-1246, 2008.

[84] T. Balomenos, A. Raouzaiou, S. Ioannou, A. Drosopoulos, K. Karpouzis and S. Kollias, "Emotion analysis in man-machine interaction systems," in *Machine Learning for Multimodal Interaction*, LNCS, Springer, 2005, pp. 318-328.

[85] H. Gunes and M. Piccardi, "Fusing face and body display for bi-modal emotion recognition: Single frame analysis and multi-frame post integration," in *Affective Computing and Intell. Interaction,* LNCS, Springer, 2005, pp. 102-111.

[86] M. F. Valstar, H. Gunes and M. Pantic, "How to distinguish posed from spontaneous smiles using geometric features," in *Int. Conf. on Multimodal Interfaces*, 2007, pp. 38-45.

[87] A. R. Abbasi, M. N. Dailey, N. V. Afzulpurkar and T. Uno, "Student mental state inference from unintentional body gestures using dynamic Bayesian networks," *J. Multimodal User Interfaces,* vol. 3, pp. 21-31, 2010.

[88] A. Kapoor and R. W. Picard, "Multimodal affect recognition in learning environments," in *ACM Int. Conf. on Multimedia,* 2005, pp. 677-682.

[89] G. Castellano et al., "Affect recognition for interactive companions: challenges and design in real world scenarios," *J. Multimodal User Interfaces,* vol. 3, pp. 89-98, 2010.

[90] P. R. De Silva, M. Osano, A. Marasinghe and A. P. Madurapperuma, "Towards recognizing emotion with affective dimensions through body gestures," in *Int. Conf. on Automatic Face and Gesture Recognition,* 2006, pp. 269-274.

[91] G. Castellano, S. D. Villalba and A. Camurri, "Recognising human emotions from body movement and gesture dynamics," in *Affective Computing and Intell. Interaction*, LNCS, Springer, 2007, pp. 71-82.

[92] L. Gong, T. Wang, C. Wang, F. Liu, F. Zhang and X. Yu, "Recognizing affect from non-stylized body motion using shape of gaussian descriptors," in *ACM Symp. on Applied Computing*, 2010, pp. 1203-1206.

[93] D. Bernhardt and P. Robinson, "Detecting emotions from connected action sequences," in *Visual Informatics: Bridging Res. and Practice,* Springer, 2009, pp. 1-11.

[94] F. E. Pollick, V. Lestou, J. Ryu and S. Cho, "Estimating the efficiency of recognizing gender and affect from biological motion," *Vision Res.,* vol. 42, pp. 2345-2355, 2002.

[95] D. Bernhardt and P. Robinson, "Detecting affect from non-stylised body motions," in *Affective Computing and Intell. Interaction,* LNCS, Springer, 2007, pp. 59-70.

[96] A. Camurri, B. Mazzarino, M. Ricchetti, R. Timmers and G. Volpe, "Multimodal analysis of expressive gesture in music and dance performances," in *Gesture-Based Commun. in Human-Computer Interaction*, LNCS, Springer, 2004, pp. 20-39.

[97] A. Kapur, A. Kapur, N. Virji-Babul, G. Tzanetakis and P. F. Driessen, "Gesture-based affective computing on motion capture data," in *Affective Computing and Intell. Interaction,* LNCS, Springer, 2005, pp. 1-7.

[98] M. Karg, K. Kuhnlenz and M. Buss, "Recognition of affect based on gait patterns," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 40, pp. 1050-1061, 2010.

[99] A. Kleinsmith, N. Bianchi-Berthouze and A. Steed, "Automatic recognition of non-acted affective postures," *IEEE Trans. Syst., Man, Cybern. B, Cybern.,* vol. 41, pp. 1027-1038, 2011.

[100] N. Savva and N. Bianchi-Berthouze, "Automatic recognition of affective body movement in a video game scenario," in *Intell. Technologies for Interactive Entertainment*, LNCS, Springer, 2012, pp. 149-159.

[101] A. Kapoor, W. Burleson and R. W. Picard, "Automatic prediction of frustration," *Int. J. Human-Comput. Studies,* vol. 65, pp. 724-736, 2007.

[102] J. Sanghvi, G. Castellano, I. Leite, A. Pereira, P. W. McOwan and A. Paiva, "Automatic analysis of affective postures and body motion to detect engagement with a game companion," in *ACM/IEEE Int. Conf. Human-Robot Interaction,* 2011, pp. 305-311.

[103] Y. Ma, H. M. Paterson and F. E. Pollick, "A motion capture library for the study of identity, gender, and emotion perception from biological motion," *Behavior Res. Methods,* vol. 38, pp. 134-141, 2006.

[104] T. Nakata, T. Mori and T. Sato, "Analysis of Impression of Robot Bodily Expression," *J. Robotics and Mechatronics,* vol. 14, pp. 27-36, 2002.

[105] M. Masuda and S. Kato, "Motion rendering system for emotion expression of human form robots based on laban movement analysis," in *IEEE Int. Symp. on Robot and Human Interactive Commun.,* 2010, pp. 324-329.

[106] M. Masuda, S. Kato and K. Itoh, "Emotion Detection from Body Motion of Human Form Robot Based on Laban Movement Analysis," *Principles of Practice in Multi-Agent Syst., LNCS,* vol. 5925/2009, pp. 322-334, 2009.

[107] K. Takahashi, M. Hosokawa and M. Hashimoto, "Remarks on designing of emotional movement for simple communication robot," in *IEEE Int. Conf. on Industrial Tech.,* 2010, pp. 585-590.

[108] T. Nomura and A. Nakao, "Comparison on Identification of Affective Body Motions by Robots Between Elder People and University Students:A Case Study in Japan," *Int. J. Soc. Robotics,* vol. 2, pp. 147-157, 2010.

[109] F. Tanaka and H. Suzuki, "Dance interaction with QRIO: A case study for non-boring interaction by using an entrainment ensemble model," in *IEEE Int. Workshop on Robot and Human Interactive Commun.,* 2004, pp. 419-424.

[110] H. Mizoguchi, T. Sato, K. Takagi, M. Nakao and Y. Hatamura, "Realization of expressive mobile robot," in *IEEE Int. Conf. on Robotics and Automation*, 1997, pp. 581-586.

[111] R. Laban, *The Mastery of Movement.* Northcote House, 1992.

[112] H. Kozima and M. Michalowski, "Keepon: A Playful Robot for Research, Therapy, and Entertainment," *Int. J. Soc. Robotics*, vol. 1, pp. 3-18, 2009.

[113] H. Song and D. S. Kwon, "Design of a robot head with arm-type antennae for emotional expression," in *Int. Conf. on Control, Automation and Syst.*, 2007, pp. 1842-1846.

[114] A. Beck, L. Cañamero and K. A. Bard, "Towards an affect space for robots to display emotional body language," in *IEEE Int. Symp. on Robot and Human Interactive Commun.,* 2010, pp. 464-469.

[115] K. Itoh et al., "Various emotional expressions with emotion expression humanoid robot WE-4RII," in *IEEE Tech. Ex. Based Conf. on Robotics and Automation,* 2004, pp. 35-36.

[116] M. Haring, N. Bee and E. Andre, "Creation and evaluation of emotion expression with body movement, sound, and eye color for humanoid robots," in *IEEE Int. Symp. on Robot and Human Interactive Commun.*, 2011, pp. 204-209.

[117] J. Li and M. Chignell, "Communication of Emotion in Social Robots through Simple Head and Arm Movements," *Int. J. Soc. Robotics,* vol. 3, pp. 125-142, 2011.

[118] M. Davis, *Guide to Movement Analysis Methods.* Technical Manuscript, 1997.

[119] M. Davis and D. Hadiks, "Non-verbal aspects of therapist attunement," *J. Clinical Psychol.*, vol. 50, pp. 393-405, 1994.

[120] M. Davis and D. Hadiks, "Nonverbal behavior and client state changes during psychotherapy," *J. Clinical Psychology,* vol. 46, pp. 340-351, 1990.

[121] D. McColl and G. Nejat, "A Human Affect Recognition System for Socially Interactive Robots," *In the Handbook of Research on Technoself: Identity in a Technological Society*, IGI Global Publishing, pp. 554-573, 2013.

[122] C. M. Bishop, *Pattern Recognition and Machine Learning.* Springer New York, 2006.

[123] M. M. Loper, N. P. Koenig, S. H. Chernova, C. V. Jones and O. C. Jenkins, "Mobile human-robot teaming with environmental tolerance," in *ACM/IEEE Int. Conf. on Human Robot Interaction*, 2009, pp. 157-164.

[124] J. Satake and J. Miura, "Robust stereo-based person detection and tracking for a person following robot," in *ICRA Workshop on People Detection and Tracking*, 2009, pp. 1-10.

[125] M. Sanders and E. McCormick, *Human Factors in Engineering and Design 7Th Edition.* New York, New York: McGraw-Hill, 1993.

[126] W. Marras and J. Kim, "Anthropometry of Industrial Populations," *J. Ergonomics*, vol. 36, pp. 371-378, 1993.

[127] G. Medioni, A. R. J. Francois, M. Siddiqui, K. Kim and H. Yoon, "Robust real-time vision for a personal service robot." *Comput. Vision and Image Understanding Archive*, vol. 18, pp. 196-203, 2007.

[128] J. Rett, J. Dias and J. Ahuactzin, "Bayesian reasoning for Laban Movement Analysis used in human-machine interaction," *Int. J. Reasoning-Based Intell. Syst.*, vol. 2, pp. 13-35, 2010.

[129] D. McColl, Z. Zhang and G. Nejat, "Human Body Pose Interpretation and Classification during Social Human-Robot Interaction," *Int. J. Soc. Robot.*, vol. 3, pp. 313-332, 2011.

[130] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann and I. H. Witten, "The WEKA data mining software: an update," *ACM SIGKDD Explorations Newsletter*, vol. 11, pp. 10-18, 2009.

[131] Y. Kim, S. Hahn and B. Zhang, "Text filtering by boosting naive bayes classifiers," in *Int. ACM SIGIR Conf. on Res. and Develop. in Information Retrieval*, 2000, pp. 168-175.

[132] G. K. Cheung, T. Kanade, J. Bouguet and M. Holler, "A real time system for robust 3D voxel reconstruction of human motions," in *IEEE Conf. Comput. Vision and Pattern Recognition*, 2000, pp. 714-720.

[133] J. R. Landis and G. G. Koch, "The measurement of observer agreement for categorical data," *Biometrics*, pp. 159-174, 1977.

[134] Thermoteknix Systems Ltd, “Thermoteknix miricle 110K,” 2011, http://www.thermoteknix.com/.

[135] CSEM, “SwissRanger SR3000,” 2007, http://www.swissranger.ch

[136] JAI Industrial Cameras, “Pulnix TMC-6740CL,” 2011 http://www.jai.com/EN/Pages/home.aspx .

[137] Bouguet, J. Y., “Matlab calibration toolbox,” 2010 http://www.vision.caltech.edu/bouguetj/calib_doc/

[138] D. Chai and K. N. Ngan, "Face Segmentation Using Skin-Color Map in Videophone Applications," *IEEE Trans. Circuits Syst. Video Tech.*, vol. 9, pp. 551-564, 1999.

[139] B. Freedman, A. Shpunt, M. Machline and Y. Arieli, "Depth Mapping Using Projected Patterns," US Patent: 20100118123 A1, May 13, 2010.

[140] D. Glowinski et al., "Toward a minimal representation of affective gestures," *IEEE Trans. Affective Computing*, vol. 2, pp. 106-118, 2011.

[141] D. R. Carney and C. R. Colvin, "The circumplex structure of affective social behavior," *Soc. Psychol. and Personality Sci.*, vol. 1, pp. 73-80, 2010.

[142] J. Posner, J. A. Russell and B. S. Peterson, "The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology," *Develop. Psychopathol.*, vol. 17, pp. 715-734, 2005.

[143] W. B. Mortenson, W. C. Miller, C. L. Backman and J. L. Oliffe, "Association Between Mobility, Participation, and Wheelchair-Related Factors in Long-Term Care Residents Who Use Wheelchairs as Their Primary Means of Mobility," *J. Am. Geriatr. Soc.*, vol. 60, pp. 1310-1315, 2012.

[144] V. Ganapathi, C. Plagemann, D. Koller and S. Thrun, "Real-time human pose tracking from range data," in *Computer Vision–ECCV 2012*, Springer, 2012, pp. 738-751.

[145] H. M. Patterson, F. E. Pollick and A. J. Sanford, "The role of velocity in affect discrimination," in *Conf. of the Cognitive Sci. Soc.*, 2001, pp. 756-761.

[146] A. Kleinsmith and N. Bianchi-Berthouze, "Recognizing affective dimensions from body posture," in *Affective Computing and Intell. Interaction*, LNCS, Springer, 2007, pp. 48-58.

[147] M. M. Gross, E. A. Crane and B. L. Fredrickson, "Effort-shape and kinematic assessment of bodily expression of emotion during gait," *Human Movement Sci.*, vol. 31, pp. 202-221, 2012.

[148] R. D. Lane, P. M. Chua and R. J. Dolan, "Common effects of emotional valence, arousal and attention on neural activation during visual processing of pictures," *Neuropsychologia*, vol. 37, pp. 989-997, 1999.

[149] M. Karg, K. Kühnlenz and M. Buss, "Recognition of Affect Based on Gait Patterns," *IEEE Trans. Syst., Man, Cybern., Part B: Cybern.*, vol. 40, pp. 1050-1061, 2010.

[150] A. B. Acharya, S. Prabhu and M. V. Muddapur, "Odontometric sex assessment from logistic regression analysis," *Int. J. Legal Med.*, vol. 125, pp. 199-204, 2011.

[151] V. Y. Kulkarni and P. K. Sinha, "Random Forest Classifiers: A Survey and Future Research Directions," *Int. J. Advanced Computing,* vol. 36, pp. 1144-1153, 2013.

[152] T. M. Mitchell, *Machine Learning.* WCB/McGraw Hill, 1997.

[153] Y. Gao, G. Ji, Z. Yang and J. Pan, "A Dynamic AdaBoost Algorithm With Adaptive Changes of Loss Function," *IEEE Trans. Syst., Man, Cybern., C: Appl. and Reviews,* vol. 42, pp. 1828-1841, 2012.

[154] H. Yu, T. Xie, S. Paszczynski and B. M. Wilamowski, "Advantages of radial basis function networks for dynamic system design," *Industrial Electronics, IEEE Transactions on,* vol. 58, pp. 5438-5450, 2011.

[155] M. R. Ade, D. S. Medhekar and M. P. Bote, "Heart Disease Prediction System Using SVM and Naive Bayes," *Int. J. Eng. Sci. & Research Tech.* vol. 2, no. 5, pp. 1343-1348.

[156] I. Rish, "An empirical study of the naive bayes classifier," in *IJCAI 2001 Workshop on Empirical Methods in Artificial Intell.,* 2001, pp. 41-46.

[157] S. Dreiseitl and L. Ohno-Machado, "Logistic regression and artificial neural network classification models: a methodology review," *J. Biomed. Inform.,* vol. 35, pp. 352-359, 2002.

[158] L. Breiman, "Random forests," *Mach. Learning,* vol. 45, pp. 5-32, 2001.

[159] D. Wettschereck, D. W. Aha and T. Mohri, "A review and empirical evaluation of feature weighting methods for a class of lazy learning algorithms," *Artif. Intell. Rev.,* vol. 11, pp. 273-314, 1997.

[160] Y. Freund and R. E. Schapire, "A desicion-theoretic generalization of on-line learning and an application to boosting," in *Computational Learning Theory*, 1995, pp. 23-37.

[161] J. Park and I. W. Sandberg, "Universal approximation using radial-basis-function networks," *Neural Comput.,* vol. 3, pp. 246-257, 1991.

[162] C. J. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining and Knowledge Discovery,* vol. 2, pp. 121-167, 1998.

[163] S. Hareli and B. Parkinson, "What's Social about Social Emotions," *J. Theory of Soc. Behaviour,* vol. 38, pp. 131-156, 2008.

[164] P. Silvia, "Interest - The Curious Emotion," *Current Directions in Psychol. Sci.,* vol. 17, pp. 57-60, 2008.

[165] J. M. Barbalet, "Boredom and social meaning," *Br. J. Sociology,* vol. 50, pp. 631-646, 2003.

[166] B. Parkinson, "Emotions are social," *Br. J. Psychol.,* vol. 87, pp. 663-683, 1996.

[167] P. R. Shaver, S. Wu and J. C. Schwartz, *Cross-Cultural Similarities and Differences in Emotion and its Representation.* Sage Publications, 1992.

[168] A. Ben-Ze'ev and K. Oatley, "The Intentional and Social Nature of Human Emotions: Reconsideration of the Distinction Between Basic and Non-basic Emotions," *J. Theory of Soc. Behaviour,* vol. 26, pp. 81-94, 1996.

[169] D. McColl and G. Nejat, "Recognizing Emotional Body Language Displayed by a Human-like Social Robot," *International Journal of Social Robotics,* vol. 6, pp. 261-280, 2014.

[170] K. C. Barrett and G. C. Nelson-Goens, "Emotion communication and development of the social emotions," in *The Commun. of Emotion: Current Res. from Diverse Perspectives,* K. C. Barrett, Ed. San Francisco: Jossey-Bass, 1997, pp. 69-88.

[171] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vision*, vol. 60, pp. 91-110, 2004.

[172] M. Heerink, B. Kröse, V. Evers and B. Wielinga, "Assessing acceptance of assistive social agent technology by older adults: the almere model," *Int. J. of Soc. Robotics,* vol. 2, pp. 361-375, 2010.

[173] J. Kehoe, "Basic Item Analysis for Multiple-Choice Tests." *Practical Assessment, Res. & Evaluation,* vol. 4, pp. 1-2, 1995.

[174] A. Libin and J. Cohen-Mansfield, "Therapeutic robocat for nursing home residents with dementia: preliminary inquiry," *Am. J. Alzheimers Dis. Other Demen.,* vol. 19, pp. 111-116, Mar-Apr, 2004.

[175] W. Taggart, S. Turkle and C. D. Kidd, "An interactive robot in a nursing home: Preliminary remarks," in *Towards Soc. Mechanisms of Android Sci.: A COGSCI Workshop,* 2005, pp. 1-11.

[176] J. Fasola and M. J. Mataric, "Using socially assistive human–robot interaction to motivate physical exercise for older adults," *Proc IEEE,* vol. 100, pp. 2512-2526, 2012.

[177] J. Shotton et al., "Real-time human pose recognition in parts from single depth images," *Commun ACM*, vol. 56, pp. 116-124, 2013.

[178] P. Refaeilzadeh, L. Tang and H. Liu, "Cross-validation," in *Encyclopedia of Database Syst.*, Springer, 2009, pp. 532-538.

[179] T. Shiwa, T. Kanda, M. Imai, H. Ishiguro and N. Hagita, "How quickly should communication robots respond?" in *ACM/IEEE Int. Conf. Human-Robot Interaction,* 2008, pp. 153-160.

[180] 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 *SIGCHI Conf. on Human Factors in Computing Syst.,* 2012, pp. 695-704.

[181] P. A. Weeks, "Error-correction techniques and sequences in instructional settings: Toward a comparative framework," *Human Studies,* vol. 8, pp. 195-233, 1985.

[182] N. Telle and H. Pfister, "Not Only the Miserable Receive Help: Empathy Promotes Prosocial Behaviour Toward the Happy," *Current Psychol.,* vol. 31, pp. 393-413, 2012.

[183] A. Comfort, *Reality and Empathy: Physics, Mind, and Science in the 21st Century.* SUNY Press, 1984.

[184] A. Bruce, I. Nourbakhsh and R. Simmons, "The role of expressiveness and attention in human-robot interaction," in *IEEE Int. Conf. on Robotics and Automation,* 2002, pp. 4138-4142.

[185] K. Hone, "Empathic agents to reduce user frustration: The effects of varying agent characteristics," *Interact. Comput.,* vol. 18, pp. 227-245, 2006.

[186] C. N. Moridis and A. A. Economides, "Affective learning: empathetic agents with emotional facial and tone of voice expressions," *IEEE Trans. Affective Computing,* vol. 3, pp. 260-272, 2012.

[187] H. Maldonado et al., "We learn better together: Enhancing eLearning with emotional characters," in *Conf. on Comput. Support for Collaborative Learning: Learning 2005: The Next 10 Years!* 2005, pp. 408-417.

[188] P. Ekman, W. V. Freisen and S. Ancoli, "Facial signs of emotional experience." *J. Pers. Soc. Psychol.,* vol. 39, pp. 1125, 1980.

[189] A. Kapoor and R. W. Picard, "A real-time head nod and shake detector," in *Workshop on Perceptive User Interfaces,* 2001, pp. 1-5.

[190] J. Montepare, E. Koff, D. Zaitchik and M. Albert, "The use of body movements and gestures as cues to emotions in younger and older adults," *J. Nonverbal Behavior,* vol. 23, pp. 133-152, 1999.

[191] O. Hargie, *Skilled Interpersonal Communication: Research, Theory and Practice.* Routledge, 2010.

[192] T. J. Housel and C. Wheeler, "The effects of nonverbal reinforcement and interviewee-interviewer relationship on interviewee's verbal response," *J. Applied Commun. Res.*, vol. 8, pp. 120-127, 1980.

[193] W. Furman and J. C. Masters, "Affective consequences of social reinforcement, punishment, and neutral behavior." *Develop. Psychol.*, vol. 16, pp. 100, 1980.

[194] D. McNeill, *Hand and Mind: What Gestures Reveal about Thought.* University of Chicago Press, 1992.

[195] D. A. Small and N. M. Verrochi, "The face of need: Facial emotion expression on charity advertisements," *J. Market. Res.*, vol. 46, pp. 777-787, 2009.

[196] N. Eisenberg, R. A. Fabes, P. A. Miller, J. Fultz, R. Shell, R. M. Mathy and R. R. Reno, "Relation of sympathy and personal distress to prosocial behavior: a multimethod study." *J. Pers. Soc. Psychol.*, vol. 57, pp. 55, 1989.

[197] M. De Boni, A. Richardson and R. Hurling, "Humour, Relationship Maintenance and Personality Matching in automated dialogue: A controlled study," *Interact. Comput.*, vol. 20, pp. 342-353, 2008.

[198] A. Niculescu, B. van Dijk, A. Nijholt, H. Li and S. L. See, "Making social robots more attractive: the effects of voice pitch, humor and empathy," *Int. J. Soc. Robotics*, vol. 5, pp. 171-191, 2013.

[199] 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, pp. 522-541, 2005.

[200] A. Field, *Discovering Statistics using SPSS.* Sage publications, 2009.

[201] 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," *Br. J. Psychiatry*, vol. 186, pp. 319-323, Apr, 2005.

[202] M. Schmid Mast, J. A. Hall, N. A. Murphy and C. R. Colvin, "Judging assertiveness," *FACTA UNIVERSITATIS-Series Philos., Sociol., Psychol. Hist.*, pp. 731-744, 2003.

[203] H. A. Elfenbein, M. K. Mandal, N. Ambady, S. Harizuka and S. Kumar, "Cross-cultural patterns in emotion recognition: highlighting design and analytical techniques." *Emotion*, vol. 2, pp. 75, 2002.

[204] B. N. Landis et al., "“Taste Strips”–A rapid, lateralized, gustatory bedside identification test based on impregnated filter papers," *J. Neurol.*, vol. 256, pp. 242-248, 2009.

[205] K. Mogg and B. P. Bradley, "Some methodological issues in assessing attentional biases for threatening faces in anxiety: A replication study using a modified version of the probe detection task," *Behav. Res. Ther.*, vol. 37, pp. 595-604, 1999.

[206] S. S. J. Neill, "Children's Reported Responses to Teachers' Non-verbal Signals: a pilot study," *J. Education for Teaching*, vol. 12, pp. 53-63, 1986.

[207] K. Aziz, "The key to perfect presentations," *Industrial and Commercial Training*, vol. 30, pp. 214-217, 1998.

[208] N. Dael, M. Mortillaro and K. R. Scherer, "Emotion expression in body action and posture." *Emotion*, vol. 12, pp. 1085, 2012.

[209] C. G. Kohler et al., "Differences in facial expressions of four universal emotions," *Psychiatry Res.*, vol. 128, pp. 235-244, 2004.

[210] C. Darwin, *The Expression of the Emotions in Man and Animals.* University of Chicago press, 1965.

[211] P. Bull, "The interpretation of posture through an alternative methodology to role play," *Br. J. Soc. Clin. Psychol.*, vol. 17, pp. 1-6, 1978.

[212] M. Bindemann, A. Mike Burton and S. R. Langton, "How do eye gaze and facial expression interact?" *Visual Cognition*, vol. 16, pp. 708-733, 2008.

[213] T. Ganel, "Revisiting the relationship between the processing of gaze direction and the processing of facial expression." *J. Experimental Psychol.: Human Perception and Performance*, vol. 37, pp. 48, 2011.

[214] J. N. Bassili, "Facial motion in the perception of faces and of emotional expression." *J. Experimental Psychol.: Human Perception and Performance*, vol. 4, pp. 373, 1978.

[215] A. Tinwell, M. Grimshaw, D. A. Nabi and A. Williams, "Facial expression of emotion and perception of the Uncanny Valley in virtual characters," *Comput. in Human Behavior*, vol. 27, pp. 741-749, 2011.

[216] C. Becker-Asano and H. Ishiguro, "Evaluating facial displays of emotion for the android robot geminoid F," in *IEEE Workshop on Affective Computational Intell.*, 2011, pp. 1-8.

[217] N. J. Roese and E. Amir, "Human—Android Interaction in the Near and Distant Future," *Perspectives on Psychol. Sci.*, vol. 4, pp. 429-434, 2009.