Simulating Social Robot Quori using ROS and Gazebo

Adrian Thinnyun[1], Roshni Kaushik[2] and Reid Simmons[2]

Abstract— **Quori is a novel and affordable socially-interactive robot platform designed for use in human-robot interaction (HRI) research. With an expressive projected face, two gesturing arms, and a bowing waist, Quori has the potential to easily generate versatile social behaviors. However, designing and implementing social skills and behaviors for Quori is currently a difficult task, due to the limited number of models available and the lack of a proper simulation environment. Having a proper simulation environment would make the development and testing of algorithms for Quori much easier. In this paper, we present our work towards a Quori simulator using ROS and Gazebo. We also showcase implementations of several behaviors for Quori that will be used in future work to determine how Quori's nonverbal movements relate to its perceived emotional state.**

Index Terms— **simulation, animation, human-robot interaction**

I. INTRODUCTION

Socially-interactive robots have experienced a surge in popularity in recent years. While traditionally robots have been most popular in industrial settings such as automobile manufacturing, recent developments in human-robot interaction (HRI) have led to a shift from industrial robotics to service robotics [1]. Service robots often share the same physical space with people and regularly interact with them in both professional settings such as storefronts and research labs and domestic applications such as assisting the elderly. In order for robots to perform well in these applications, they must have strong capabilities for social interaction.

Two popular socially-interactive robots are Softbank's Pepper and Nao. Pepper and Nao are both humanoid robots that have been used in a wide variety of professional, research, and recreational activities. Pepper can be found in several Softbank stores around Japan waiting to greet customers and has proven to be equally popular with adults and children alike [2]. Nao has become a popular contender in the annual RoboCup league, which features intelligent autonomous robots competing in soccer [3]. Although both robots are widely used in labs and universities across the world, some research groups may not be able to afford the just-under \$2000 price tag of Pepper [4], or Nao's nearly \$8000 price [5]. Thus, having an affordable alternative may increase the opportunities for further HRI research.

Quori was designed to fulfill this role, released in the summer of 2018 with ten platforms awarded to ten research groups [6]. It was developed with input from the HRI community to prioritize the most important hardware capabilities for social interaction and maximize functionality while keeping costs affordable. For example, Quori's head consists of a low-cost projector allowing it to display various faces/expressions easily, and its arms are connected by 2-DOF shoulder joints that allow for simple gestures. A 1-DOF spine attaches this upper body to a holonomic mobile base, and options are available for mounting cameras, microphones, and other sensors to extend Quori's capabilities. Fig. 1 illustrates a sample configuration of Quori, though note that the modularity of its design allows for other configurations.

We use the open-source software Gazebo [8] as the environment for our simulator. Gazebo is a popular robot simulator that has already been used to simulate robots frequently used in HRI research, including both Pepper and Nao. A Unified Robot Description Format (URDF) [9] file is used to describe the virtual model of Quori, containing information about the various links of the model and their associated masses, inertia matrices, and joints. This information is extracted by the engine and used to render the robot in a 3D environment, as well as to simulate physical forces and enforce collision checking. We also use the ROS [10] framework in order to interact with the virtual robot model, sending actions for the robot to perform by publishing messages to the appropriate control topics.

In addition to developing a simulator for Quori, we also present work on developing non-verbal behaviors appropriate for social interaction. Humans use a wide variety of poses and gestures to convey emotions and other affective infor-

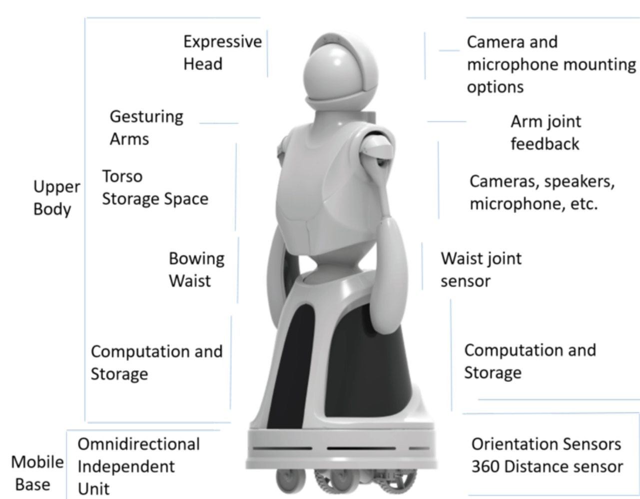

Fig. 1. Diagram of Quori illustrating its components and available degrees of freedom [7]

[1]Adrian Thinnyun is with the Department of Computer Science, University of Virginia, Charlottesville, USA athinnyun@gmail.com

[2]Roshni Kaushik and Reid Simmons are with the Robotics Institute, Carnegie Mellon University, Pittsburgh, USA
{roshnika, rsimmons}@andrew.cmu.edu

mation over the course of natural conversation. For example, someone experiencing anxiety or panic may lean backwards, while someone experiencing joy may move their arms up and down [11]. In turn, seeing someone express a certain emotion non-verbally often provokes an emotional response in oneself. Behaviors associated with disgust often create responses of fear, while behaviors associated with anger provoke more anger [12]. Thus, having the ability to produce these gestures seamlessly and accurately within the context of a conversation is critical to Quori's capability for social interaction.

One application of Quori's social behaviors we are particularly interested in is education. Recently there has been a promising body of work showing the effectiveness of the "learning by teaching" paradigm involving students instructing teachable robots [13]–[16]. Tanaka et al. used Pepper to teach children English through a variety of programs. In one activity, Pepper uses its tablet to display a lesson teaching the word "plane" and then extends its arms outwards to imitate the wings of a plane, inviting the children to join as they learn the word. In another activity, Pepper shows a video of a teacher teaching the word "mouth" while placing Pepper's hand on its mouth, then tells the children, "Teach me like that teacher is doing". This method, known as "direct teaching", has been shown to be effective in promoting children's learning [17].

Like Pepper, Nao has also seen use in experiments focused on improving education. Lemaignan et al. used Nao to help young children with motor difficulties improve their handwriting. Rather than Nao teaching the children how to write, the children were tasked with teaching Nao how to write, with the help of an occupational therapist. Although these children experience great deals of anxiety dealing with their handwriting difficulties, this scenario transforms the role of the child from a "bad writer" to a "teacher" helping Nao with its own handwriting, a strategy that greatly improves the child's self-esteem and motivation to participate in the activity. These studies demonstrate the potential of robots and the "learning by teaching" paradigm.

Quori could be a potential candidate for research of this nature, but before it can be used in any experiments it needs to be able to socially interact with humans by understanding the emotional state of its partner and responding with appropriate, context-driven actions. Previous work has shown a strong relationship between a student's nonverbal behaviors and their teacher's expectations and self-assessments [18], so care must be taken in order to make sure that Quori responds appropriately to a student mentor's instruction. In this paper, we present our progress on a Quori simulator and work towards developing a diverse set of nonverbal social behaviors.

II. SIMULATION OVERVIEW

In this section, we give a brief overview of the core components of our Quori simulation and present our original contributions to it.

A. Universal Robot Description Format (URDF)

At the heart of the simulation lies Quori's URDF file, which contains all of the necessary information for recreating Quori in a virtual environment. The most basic URDF file simply contains all of the links and joints comprising the robot defined in XML, where links are visually represented using either basic geometric shapes or pre-defined mesh files. However, in order to properly simulate a robot in Gazebo's physics engine, one must also add inertial information, including the mass and moment of inertia. Additional tags specifying colors and textures for each link, while not required, are helpful for accurately recreating the robot's true appearance and we have included this information in our URDF file. Finally, we include transmission tags describing the actuators of Quori's movable joints, as these allow us to manipulate our simulated robot using `ros_control`.

B. Robot Control in ROS

The `ros_control` [19] framework allows for control of the robot using standard ROS topics, nodes, and messages. A variety of controllers are available, each with its own parameters, message type, and function. We chose to use a `joint_trajectory_controller` as the main controller for Quori's joints. This controller accepts `JointTrajectory` messages as input, which contain a list of waypoints specifying the position, velocity, and acceleration of each joint at various time steps. We determined that this controller type provided the necessary freedom and flexibility to accurately produce the diverse set of movements needed for nonverbal communication.

C. Inverse Kinematics

To make the description of arm movements easier, we have also implemented an inverse kinematics solution that converts the 3-D desired position of the arm's end-point into the two shoulder angles required to reach the desired point. The equations for this solution were provided to us by the developers of Quori at the University of Pennsylvania.

For a given arm, we define θ_1 as the rotation of the arm's shoulder joint and θ_2 as its abduction/adduction (i.e. the raising or lowering of the arm). As depicted in Fig. 2, C_0 indicates the origin of the arm joint and p_a denotes an arbitrary point on the arm. Finally, we define the coordinate plane such that the x-axis points out to the front of the robot, the y-axis points out to the side of the robot from which the arm originates, and the z-axis is normal to the ground plane (assuming the torso is vertical). We chose to define the coordinate planes for the left and right arms as reflections of one another so that inputting the same command for both arms would lead to symmetrical movement. Thus, for an end-point of the arm $p_a = (x, y, z)$, the equations for the two joint angles are:

$$\theta_1 = \mathrm{atan2}(z/x) \tag{1}$$

$$\theta_2 = \mathrm{atan2}\left(\frac{\sqrt{x^2+z^2}}{-y}\right) \tag{2}$$

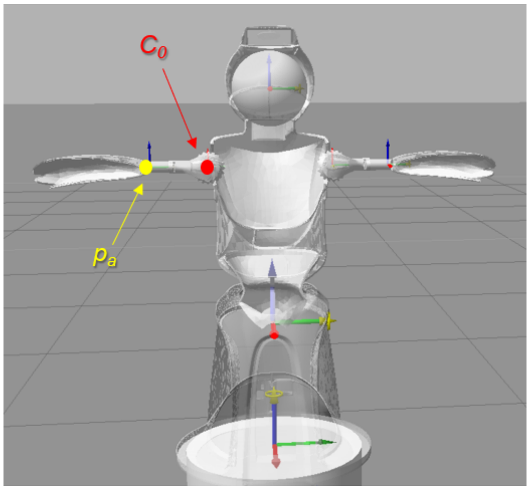

Fig. 2. Visualization of Quori, where C_0 indicates the origin of the arm joint and p_a denotes an arbitrary point on the arm (in this case, the origin of the lower arm)

This solution is implemented by a ROS node which subscribes to a topic consisting of `Float64MultiArray` messages specifying the arm(s) to be moved, the desired time for the action, and the Cartesian coordinates to be reached. The node converts these messages to `JointTrajectory` messages using Equations (1) and (2) and publishes them to the `joint_trajectory_controller`'s command topic.

III. NONVERBAL BEHAVIOR DESIGN

In this section, we discuss the key considerations and concepts involved in designing nonverbal social behaviors for Quori.

A. Facial and Bodily Expressions

With the necessary mechanisms in place for executing complex movements with Quori in a simulated environment, we focus our discussion on the human nonverbal behaviors we wish to emulate. In particular, we focus on bodily expressions of emotion rather than facial expressions. Originally it was believed that facial and vocal expressions alone were responsible for communicating the type of emotion being experienced while bodily movements/postures merely conveyed the intensity or level of arousal of the emotional state [20], [21]. However, this view has been challenged by recent studies showing that dynamic whole body and arm movement [22]–[25] as well as static body postures [26], [27] reveal specific information about a person's emotional state. Thus it is imperative to take Quori's whole body configuration into consideration when designing effective affective non-verbal behaviors.

B. The Action-Perception Disconnect

While the literature has found that certain body movements/poses seem to be associated with certain emotional states, performing those behaviors does not necessarily entail that an observer will perceive the correct emotional state from them. Perceived emotion may vary greatly depending on context, including factors such as the relationship between the actor and the observer, the environment in which the interaction takes place, the emotional state of the observer, etc. Thus, when designing nonverbal behaviors, we must acknowledge that we are merely optimizing behaviors towards the goal of successfully conveying a desired affect, rather than creating behaviors that do convey a desired affect.

With Quori, we encounter an additional disconnect: the disconnect between human and robot. Even if a certain human behavior always lead to a certain perceived emotional state and we were able to replicate it perfectly with Quori, we still would not be able to guarantee that an observer would interpret the behavior the same as they would if a human had performed it instead. Some people do not believe robots can have emotions, or that robots could take on the same roles as a human (e.g. student, teacher, salesperson, etc.). The visual differences between a robot and a human could also affect its perceived affect. Thus any behaviors we create for Quori will require further user validation before they can be deployed with confidence.

C. Quori's Limitations

Since Quori is a humanoid robot, it makes sense to use human body movements as a starting point when designing nonverbal behaviors for it, even if they may not be interpreted in the same manner. However, translating human body movements and gestures into motions Quori can perform is a difficult task. Consider the array of emotions that can be expressed by a single human hand. One can raise their fist to the sky, shake it back and forth, extend an open hand in front of them, or gently hold it above their head. Each of these actions could be used to make different inferences about the actor's intentions/emotions depending on the context and precise motions used. Quori lacks the ability to perform any of these actions as it has neither fingers nor a wrist. Similarly, simple gestures like nodding in agreement or shaking one's head in disagreement are unavailable due to Quori's lack of a movable neck joint.

Another limitation to consider is the speed at which Quori is able to move its joints. Humans often use quick, abrupt movements when expressing emotion, such as a sudden fist pump when feeling joy or quickly outstretching one's arms when feeling panic or fear. Quori's motors are only capable of producing relatively smooth, continuous motion, so these kinds of gestures are also unavailable to Quori. Quori's available degrees of freedom (DOF) are summarized in Table I.

TABLE I
QUORI'S DEGREES OF FREEDOM

Joint Name	Upper Limit	Lower Limit	Speed
Waist Hinge	0.47 radians	-0.27 radians	1 rad/sec
Turret Joint	N/A	N/A	3 rad/sec
Arm (Rotation)	N/A	N/A	1 rad/sec
Arm (Abduction)	1.1 radians	-1.1 radians	1 rad/sec

D. Mapping Body Movements to DOF

We now present our initial design choices for Quori's nonverbal behaviors. Table II features the six canonical emotions – happiness, sadness, fear, disgust, anger, and surprise – as well as the state of "interest", which we believe will be useful in an educational setting. For each affect, the representative postures and movements found in the literature are listed along with the analogous behaviors that can be performed by Quori.

At a glance each behavior seems reasonable on its own, but a problem arises when considering the resulting set of Quori behaviors in conjunction with one another. There are several cases of multiple affects sharing similar or identical poses. For example, both happiness and surprise are represented by the torso leaning backward and arms raised high. Similarly, both sadness and interest are represented by the torso leaning forward and arms resting by Quori's side. Thus, these poses currently lead to emotional ambiguity, which is unhelpful for social interaction. We discuss steps for resolving this issue in our section on future work.

IV. EXPERIMENTAL RESULTS

A. Preliminary Gestures

Using our simulated model as a basis, we have implemented a few nonverbal behaviors to showcase Quori's capabilities for social interaction. As seen in Fig. 3, the first of these involves Quori slowly leaning backward while raising its arms above its head, while the second involves Quori quickly leaning forward and stretching its left hand forward. In Table II, the former of these motions may be associated with happiness, while the latter most closely corresponds to anger, but we again stress that the actual perceived emotional state arising from these actions is unknown without user validation. In addition to these two simple behaviors, we have also implemented a more dynamic variant of the first expression, which consists of Quori raising its arms above its head and using its turret joint to quickly spin around 360 degrees. This behavior illustrates two important concepts. First, while there are many motions that can be performed by a human but not Quori, there are some motions that can be easily performed by Quori but not by a human. Secondly, the novelty of dynamic animations and movements performed by a robot could prove effective in provoking interest and fascination from a student, strengthening their willingness to engage with Quori and, consequently, their own learning. However, once again we will require additional work in order to see whether this effect is observed in a real-life setting.

V. FUTURE WORK

In the future, we would like to expand on this work in a few key ways. As mentioned previously, there currently exists some ambiguity in the behaviors we have derived from the literature. Differentiating the expressions of each emotional state from one another as much as possible is important for accurate conveyance of Quori's desired affect. Additionally, we need user validation to determine whether the intended perceived emotional state of each behavior corresponds to its actual perceived emotional state. We intend to conduct a survey asking participants to categorize videos of Quori performing various poses and motions and report the affect they perceive. These videos would include additional variability in the movements by including various combinations of start/end positions, speed, degrees, and asymmetrical motion. The results of this study will not only help us differentiate our behaviors but also validate their effectiveness in conveying emotional states.

Additionally, although we have precise control over Quori's upper-body motions, our control of Quori's wheels and movement is currently limited. We plan to implement a differential drive controller based on the Jacobian to coordinate the base and turret motors to achieve omnidirectional capability. This would allow us to begin developing coordinated motions such as facing a given point in space while moving. Combined with the implementation of cameras, laser scanners, and other sensors for our simulated model, these developments would greatly enhance Quori's capabilities.

Lastly, we would like to implement these behaviors on a real Quori robot and see how they function in a real-world setting. Our Quori has not yet been delivered, so we cannot currently test using real hardware. If effective, these behaviors would be instrumental towards preparing Quori to interact with students in an educational setting. Having been designed with affordability in mind, Quori could one day populate many classrooms to aid in student education.

ACKNOWLEDGMENT

This material is based upon work supported by the National Science Foundation under Grant No. 1659774. I would like to thank Dr. Reid Simmons and Roshni Kaushik for their mentorship, encouragement, and support on this project throughout the summer. I would also like to thank Rachel Burcin and John Dolan for supporting me and my peers over the course of the RISS program.

REFERENCES

[1] S. Thrun, "Toward a framework for human-robot interaction," *Human–Computer Interaction*, vol. 19, no. 1-2, pp. 9–24, 2004.

[2] I. Aaltonen, A. Arvola, P. Heikkilä, and H. Lammi, "Hello pepper, may i tickle you? children's and adults' responses to an entertainment robot at a shopping mall," in *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, 2017, pp. 53–54.

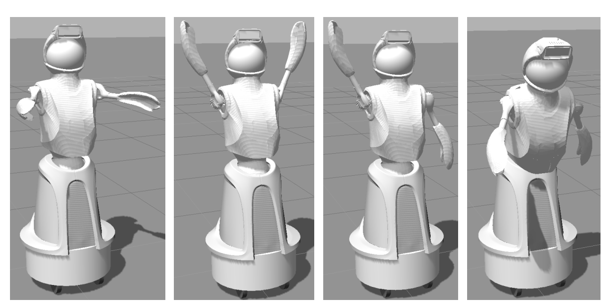

Fig. 3. Non-verbal behaviors performed by Quori in Gazebo

TABLE II
MAPPING MOVEMENTS TO QUORI DOF

Emotion	From Literature	Quori Translation
Happiness	Symmetrical up-down motion of the arms [11] Hands kept high; hands made into fists and kept high [12] Slight lean backwards, arms raised high [26]	Torso leans backward Arms raised high, Arms moving up and down symmetrically
Sadness	Leaning forward, Hands at sides [12], [26] Hands over head [28]	Torso leans forward Arms at sides
Fear	Leaning backward [11] Hands out to sides [12] Body backing, Hands over head, trying to cover body parts [28] (Anxiety) Fingers moving, fingers tapping [28] Leaning backward, Arms slightly forward [26]	Torso leans backward Arms out to sides
Disgust	Leaning backward, Arms forward [26]	Torso leans backward Arms forward
Anger	Leaning forward [11], [12] Arms crossed/on hips [12] Hands on waist, hands into fist or kept low, high speed lift of the hand [28] Leaning forward, Arms forward [26]	Torso leans forward Arms forward
Surprise	Hands over head [28] Leaning backward, Hands over head [26]	Torso leans backward Arms raised high
Interest	Leaning forward, Arms resting at side [11]	Torso leans forward Arms resting at side

[3] D. Albani, A. Youssef, V. Suriani, D. Nardi, and D. D. Bloisi, "A deep learning approach for object recognition with nao soccer robots," in *Robot World Cup*. Springer, 2016, pp. 392–403.

[4] F. Tobe, "How is pepper, softbank's emotional robot, doing?" https://www.therobotreport.com/how-is-pepper-softbanks-emotional-robot-doing/, 2016, accessed: 2020-08-04.

[5] A. Moon, "Nao next gen now available for a wider audience," https://robohub.org/nao-next-gen-now-available-for-the-consumer-market/, 2014, accessed: 2020-08-04.

[6] A. Specian, N. Eckenstein, M. Yim, R. Mead, B. McDorman, S. Kim, and M. Matarić, "Preliminary system and hardware design for quori, a low-cost, modular, socially interactive robot," in *Workshop on social robots in the wild*, 2018.

[7] A. Specian, "About quori," https://www.quori.org/about, 2018, accessed: 2020-07-30.

[8] N. Koenig and A. Howard, "Design and use paradigms for gazebo, an open-source multi-robot simulator," in *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)(IEEE Cat. No. 04CH37566)*, vol. 3. IEEE, 2004, pp. 2149–2154.

[9] W. Garage, "Universal robot description format (urdf)," http://wiki.ros.org/urdf/, 2009, accessed: 2020-07-30.

[10] M. Quigley, K. Conley, B. Gerkey, J. Faust, T. Foote, J. Leibs, R. Wheeler, and A. Y. Ng, "Ros: an open-source robot operating system," in *ICRA workshop on open source software*, vol. 3, no. 3.2. Kobe, Japan, 2009, p. 5.

[11] N. Dael, M. Mortillaro, and K. R. Scherer, "Emotion expression in body action and posture." *Emotion*, vol. 12, no. 5, p. 1085, 2012.

[12] S. Buisine, M. Courgeon, A. Charles, C. Clavel, J.-C. Martin, N. Tan, and O. Grynszpan, "The role of body postures in the recognition of emotions in contextually rich scenarios," *International Journal of Human-Computer Interaction*, vol. 30, no. 1, pp. 52–62, 2014.

[13] F. Tanaka, K. Isshiki, F. Takahashi, M. Uekusa, R. Sei, and K. Hayashi, "Pepper learns together with children: Development of an educational application," in *2015 IEEE-RAS 15th International Conference on Humanoid Robots (Humanoids)*. IEEE, 2015, pp. 270–275.

[14] S. Lemaignan, A. Jacq, D. Hood, F. Garcia, A. Paiva, and P. Dillenbourg, "Learning by teaching a robot: The case of handwriting," *IEEE Robotics & Automation Magazine*, vol. 23, no. 2, pp. 56–66, 2016.

[15] N. Lubold, E. Walker, H. Pon-Barry, Y. Flores, and A. Ogan, "Using iterative design to create efficacy-building social experiences with a teachable robot." International Society of the Learning Sciences, Inc.[ISLS]., 2018.

[16] L. Pareto, T. Arvemo, Y. Dahl, M. Haake, and A. Gulz, "A teachable-agent arithmetic game's effects on mathematics understanding, attitude and self-efficacy," in *International Conference on Artificial Intelligence in Education*. Springer, 2011, pp. 247–255.

[17] F. Tanaka and S. Matsuzoe, "Children teach a care-receiving robot to promote their learning: Field experiments in a classroom for vocabulary learning," *Journal of Human-Robot Interaction*, vol. 1, no. 1, pp. 78–95, 2012.

[18] T. P. Mottet and V. P. Richmond, "Student nonverbal communication and its influence on teachers and teaching," *Communication for teachers*, pp. 47–61, 2002.

[19] S. Chitta, E. Marder-Eppstein, W. Meeussen, V. Pradeep, A. Rodríguez Tsouroukdissian, J. Bohren, D. Coleman, B. Magyar, G. Raiola, M. Lüdtke, and E. Fernández Perdomo, "ros_control: A generic and simple control framework for ros," *The Journal of Open Source Software*, 2017. [Online]. Available: http://www.theoj.org/joss-papers/joss.00456/10.21105.joss.00456.pdf

[20] P. Ekman, "Differential communication of affect by head and body cues." *Journal of personality and social psychology*, vol. 2, no. 5, p. 726, 1965.

[21] P. Ekman and W. V. Friesen, "Head and body cues in the judgment of emotion: A reformulation," *Perceptual and motor skills*, vol. 24, no. 3 PT 1, pp. 711–724, 1967.

[22] A. P. Atkinson, W. H. Dittrich, A. J. Gemmell, and A. W. Young, "Emotion perception from dynamic and static body expressions in point-light and full-light displays," *Perception*, vol. 33, no. 6, pp. 717–746, 2004.

[23] A. P. Atkinson, M. L. Tunstall, and W. H. Dittrich, "Evidence for distinct contributions of form and motion information to the recognition of emotions from body gestures," *Cognition*, vol. 104, no. 1, pp. 59–72, 2007.

[24] 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," *Journal of Nonverbal Behavior*, vol. 23, no. 2, pp. 133–152, 1999.

[25] F. E. Pollick, H. M. Paterson, A. Bruderlin, and A. J. Sanford, "Perceiving affect from arm movement," *Cognition*, vol. 82, no. 2, pp. B51–B61, 2001.

[26] M. Coulson, "Attributing emotion to static body postures: Recognition accuracy, confusions, and viewpoint dependence," *Journal of nonverbal behavior*, vol. 28, no. 2, pp. 117–139, 2004.

[27] J. L. Tracy and R. W. Robins, "Show your pride: Evidence for a discrete emotion expression," *Psychological Science*, vol. 15, no. 3, pp. 194–197, 2004.

[28] I.-O. Stathopoulou and G. A. Tsihrintzis, "Emotion recognition from body movements and gestures," in *Intelligent Interactive Multimedia Systems and Services*. Springer, 2011, pp. 295–303.