Multi-Modal Perception and Behavior Adaptation Models for Human State Understanding and Interaction Improvement in Robotic Touch

Huy Quyen Ngo[1,*], **Rana Soltani Zarrin**[2]
[1] The Robotics Institute, Carnegie Mellon University
[2] Honda Research Institute, USA
huyquyen@andrew.cmu.edu, rana_soltanizarrin@honda-ri.com

Abstract

Robots that can physically interact with humans in a safe, comfortable, and intuitive manner can help in a variety of settings. However, perceptions of the users greatly affect the acceptability of such robots. Ability of the system to understand user's perception of the physical interaction as well as adapting robot's behaviors based on user perception and interaction context can facilitate acceptability of these robots. In this paper we propose a perception-based interaction adaptation framework. One main component of this framework is a multi-modal perception model which is grounded on the existing literature and is intended to provide a quantitative estimation of the human state- defined as the perceptions of the physical interaction- by using human, robot, and context information. This model is intended to be comprehensive in many physical Human-Robot Interaction (pHRI) scenarios. The estimated human state is fed to a context-aware behavior adaptation framework which recommends robot behaviors to improve human state using a learned behavior cost model and an optimization formulation. We show the potential and feasibility of such a human state estimation model by evaluating a reduced model, with data collected through a user study. Additionally, through some feature analysis, we aimed to shed light on future interaction designs for pHRI.

Introduction

Robots may make physical interaction with humans in a variety of contexts, from instrumental for medical reasons (Chen et al. 2014), to affective for social reasons (Zheng et al. 2019). How humans perceive those interactions would affect the level of integration of such robots in our societies. In existing literature, human perception when physically interacting with robots, namely perceived safety (Rubagotti et al. 2022), comfort and trust (Akalin, Kristoffersson, and Loutfi 2022), were investigated under different settings (i.e., robot giving bed bath, assisted feeding) and using different robotic platforms (i.e., robot arm, mobile robot) (King et al. 2010; Hu et al. 2020). However, previous works focused on studying human perception in a single context (e.g. task, body part), with one particular robot and around specific robot parameters and human backgrounds, and the physical interaction did not necessarily involve robot-initiated touch. In addition to the qualitative nature of such revealed interconnections, past work did not intend to develop a model of human perception of physical interaction based on a comprehensive collection of attributing factors, such that it may be applied to robotic touch settings within a variety of contexts, robot platforms and parameters, and human backgrounds, and it can quantitatively output the human state. With the purpose of enabling robots to perform safe and human friendly interactions, in this work we aim to develop a multi-modal human perception model for physical human-robot interaction (pHRI) that is comprehensive to be applicable to most robot-initiated touch scenarios. By considering variables such as the inherent robot attributes human background, robot behaviors, human responses, and interaction context, the model is intended to be utilizable in various pHRI scenarios. While evaluation of the full model requires extensive data collection through a separate study, in the current work we show the feasibility of such a model by evaluating a reduced model through the data collected for some representative robot-initiated touch scenarios.

While understanding human perception is the first step toward a more pleasant interaction, adapting the robot behaviors in accordance with the perceived state of human or their preference is the next necessary step, however, the literature lacks such developments. Thus, in this work, we designed the framework of a context-aware behavior adaptation model for robot-initiated touch with verbal and nonverbal behaviors to improve human state during interaction. In the core of the system, we introduced the behavior cost model, through which the robot can learn the communication preference of humans based on human background, human state, and interaction context to recommend a set of behaviors that could potentially improve the state of human in terms of perceived safety, comfort, and intuitiveness.

In summary, our contributions in this paper include:

1. A human perception model for estimating human state that can be applied to most pHRI scenarios, specifically robot-initiated touch.
2. A framework on robot behavior adaptation with verbal and nonverbal behaviors for improving human states.
3. Guideline on designing pHRI behaviors for improving perceived human safety, trust, and comfort in pHRI.

---

[*] Work done while interning at Honda Research Institute USA.

Related Work

Human Perception of Robot-Initiated Touch

Previous work have studied some of the parameters that may influence human's perception during interaction with a robot. Akalin (Akalin, Kristoffersson, and Loutfi 2022) claimed that perceived safety in human-robot interaction (HRI) is associated with comfort and trust, and that human background such as personality, gender, age, and robot familiarity have influence on this perception. Hu (Hu et al. 2022) concluded that human posture and robot proximity have effect on human's physical and mental state during an active physical interaction with a robot. And Hoffmann (Hoffmann and Krämer 2021) suggested that motion parameters like speed, delay, force, and softness of the touch contribute to human perception and emotions while interacting with humanoid robot NAO (Gouaillier et al. 2009). In addition, parameters such as robot size and appearance (Rubagotti et al. 2022), shape, color, and mobility mechanism (Kanda et al. 2008), and level of robot autonomy (Gutman, Olatunji, and Edan 2021) are identified as additional influencing factors. Except for such scattered studies, HRI literature is lacking a comprehensive model that can determine human state based on the collective influencing factors.

Unlike general HRI, human's perception during robot-initiated touch interactions is a much less explored area. Chen (Chen et al. 2014) investigated the task of robot wiping participants' forearms, in which participants have more positive responses to instrumental touch than affective touch. King (King et al. 2010) claimed that participants feel more trust towards the robot if they have more exposure to the robot, and their perception change with the touch location. Similarly Zheng (Zheng et al. 2019) concluded that touch position, length, and force can cause changes in human emotions. Except for such qualitative explorations, there is currently no human perception model for determining the human state in robot-initiated touch scenarios within multiple contexts and considering various influencing parameters. Our work addresses this shortcoming through a multi-modal model that can quantify human states during physical human-robot interaction using robot, human, and context information, and such that the model can be used across various settings. While our human perception model is a collection of influencing parameters mainly grounded on the existing pHRI literature, we added additional modalities which we believe can affect human perception.

Behavior Adaptation for Improved Interaction

To accommodate human preferences, previous works have explored robot behavior modifications in social robotics through some parameter adjustments. Mitsunaga (Mitsunaga et al. 2008) proposed a policy gradient reinforcement learning (RL) method to adapt robot's behaviors such as distance to human, robot gaze direction, timing and speed of gestures based on the human's comfort level. Later works used RL for the problem of robot parameter adaptation for making robot's behavior less boring and more effective for long-term human-robot interaction (Magyar and Vircikova 2015). However, adaptation of robot behaviors including communication behaviors to improve human perception in pHRI, specially in the context of instrumental robot-initiated touch, has not been explored before. To have a method that can be applied to various contexts of pHRI in an online manner without need for large human interaction data for RL training, we propose an adaptation framework which considering information about human like background and state and task context, and using a regression-based behavior cost prediction model, recommends best feasible adaptation behaviors. We formulated recommendation framework as a knapsack problem which has been used in some recent work in recommendation systems (Ozsoydan and Gölcük 2023; Thongsri et al. 2022), however, knapsack has not been studied in the pHRI context before.

Perception-based Interaction Adaptation

The overview of the proposed system is shown in Fig. 1. Data from human (background, human responses), context (task, level of emergency), and robot (robot attributes and robot behavior) are fed into the perception model, which outputs the estimated human state (perceived safety, perceived comfort, perceived trust). For the Behavior Adaptation, data from human (background), context (task and level of emergency), current human state, along with budget, namely execution time, are passed in. The behavior adaptation module contains a regression model that learns the costs and benefits associated with each behavior based on the human background, context, and current human state. Then, those estimated costs and benefits of the behaviors along with a budget value are passed to a Knapsack-based optimization solver which is intended to generate the recommendations for communication behaviors and robot parameter adaptations. The recommended behaviors are then executed and fed back into the Perception Model for the next iteration.

Perception Model

The purpose of the perception model is to estimate human state during physical human-robot touch interactions based on human, robot, and context inputs. This model is presented in Fig. 2. Utilizing the literature, we proposed five sub-modules in the perception model: *Robot Attributes*, *Robot Behavior*, *Human Background*, *Human Responses*, and *Context*, which will be individually described in the following subsections. In addition to information extracted from literature, we proposed additional factors and more in-depth customized categorizations of factors, namely "Contact Conformity", "Physical Condition", "Degrees of Freedom", "Level of Emergency", as well as categorizations of "Mobility", "Appearance", and "Level of Autonomy"' which we believe can also affect human perception. For more real-world robotic usage scenarios, we developed the model based on sensory information that can be obtained by the robot and does not need human wearable sensors (thus the exclusion of physiological parameters such as heart rate). Current model inputs are obtainable through computer vision, brief pre-screen questions, and internal parameters of the robot.

Robot Attributes This module contains the inherent parameters of the robot that are unchanged in the course

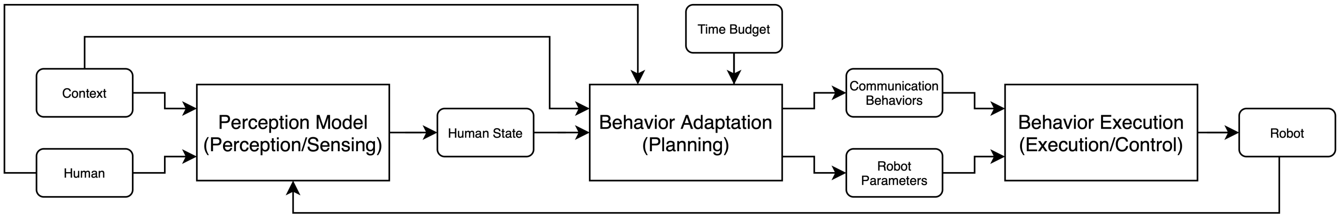

Figure 1: Overview of the Perception-based Interaction Adaptation System.

of physical interaction. Those parameters include *size* (Rubagotti et al. 2022) (the physical size of the robot), *mobility* (Kanda et al. 2008) (robot's possessed mode of mobility, i.e., legs, wheels, none), *appearance* (Rubagotti et al. 2022) (physical appearance of the robot, i.e., anthropomorphic, zoomorphic, mechanoid), *shape* (Rubagotti et al. 2022) (whether the robot has curved or sharp surfaces), *DoF* (degrees of freedom the robot has), and *skin* (the color (Kanda et al. 2008), robot's body material and end-effector contact material (Hoffmann and Krämer 2021), the part that comes into contact with humans, such as rubber glove for robot hand). It also includes *level of autonomy* (Gutman, Olatunji, and Edan 2021) of the robot (adapted from 10 (Beer, Fisk, and Rogers 2014) to 4 levels), which are No Autonomy (human-operated), Low Autonomy (all robot parameters are input by users), Some Autonomy (the robot plans its actions but asks users for permission to execute), and High Autonomy (the robot plans and acts autonomously).

Robot Behavior This module contains parameters of the robot that could be changed during physical interaction and each can be mapped to a discretized level in their defined range. Those parameters include *contact conformity* (a proposed parameter representing how conformed the robot hand is to human body part), *force* (Hoffmann and Krämer 2021) (force applied by the robot hand), *speed* (Hoffmann and Krämer 2021) (how fast the robot hand moves), *body part* (King et al. 2010) (the interaction location on human's body), *trajectory* (Rubagotti et al. 2022; Chen et al. 2014; Ngo and Steinfeld 2024) (timings of motions and the number of end-effector waypoints), *proximity* (Hu et al. 2022) (how close the robot is to human), and *communication* (binary verbal and nonverbal communication such as gaze (Hirano et al. 2018; Shiomi et al. 2020), facial expression (Peña and Tanaka 2020), nonverbal gestures (Ngo, Carter, and Steinfeld 2024), and speech (Mazursky, DeVoe, and Sebo 2022)).

Human Background This module contains the information about the human user who interacts with the robot, mostly inspired by (Akalin, Kristoffersson, and Loutfi 2022). It include *age*, *gender* (categorized), *ethnicity* (categorized), and *robot familiarity* (level of familiarity with or previous exposure to the robot). Finally, *physical condition* of the human user is a proposed parameter.

Human Responses This module represents existence of verbal and physical responses from humans when physically interacting with robots. Those responses include *speech* (i.e., (Alonso-Martin et al. 2013)) (whether and what humans verbally utter) and *pitch* (i.e., (Yutaka and Ken 2005)) (human's verbal utterance pitch), *facial expression* (Hoffmann and Krämer 2021), *gaze* (Hoffmann and Krämer 2021; Hu et al. 2020; Rubagotti et al. 2022) (human's gaze direction), and *proximity* (Hu et al. 2022) (human's distance from robot).

Context This module contains the *task context* (Chen et al. 2011; Mazursky, DeVoe, and Sebo 2022; Chen et al. 2014) (whether the interaction has affective or instrumental nature) and *level of emergency* (proposed parameter) of the situation that the human is in (i.e., medical situations).

Behavior Adaptation Model

We assume that robot actions -defined as robot behaviors-which could be verbal or non-verbal and communication or non-communication related, can change the state of the human. However, to fully consider human preference of actions, their perception on how time consuming or interruptive an action could be needs to be considered as well, when applicable. We assume such preferences can be learned through data collected from humans, either through questionnaires or experiments. Thus, in order to improve human's perception of interaction, the adaptation module recommends a set of behaviors that can be performed within a certain time and can maximize the total sum of benefits (representing benefit-cost). This representation fits well with the Knapsack problem formulation (Martello and Toth 1987).

Problem Formulation We assume that the robot can perform N behaviors in total, where each behavior, denoted by i, has some cost c_i and benefit b_i associated with it and takes time t_i to be performed. There exists a maximum cost T as the total time budget for the entire behaviors to be performed (does not include human's response time during the sequence of robot's behaviors). Additionally, there may be Mutual Exclusivity M between pairs of behaviors which prevent simultaneous execution of those behaviors (i.e., *Gaze At User* and *Gaze At Touch Location*). $M = [m_{i,j}]$ is an $N \times N$ symmetric, binary matrix with zero diagonal, where

$$m_{i,j} = \begin{cases} 1 & \text{if } i \text{ and } j \text{ are mutually exclusive} \\ 0 & \text{otherwise.} \end{cases}$$

We model our behavior adaptation as a modified 0-1 Knapsack problem from (Fréville 2004), with the introduction of Mutual Exclusivity M (Eq. 1). The optimization determines the behaviors to be performed by the robot (behavior i will be performed if $x_i = 1$ and not if $x_i = 0$) such that the weighted sum of the benefits and costs associated with the behaviors is maximum, while the total execution time of all the selected behaviors stays below T (design choice).

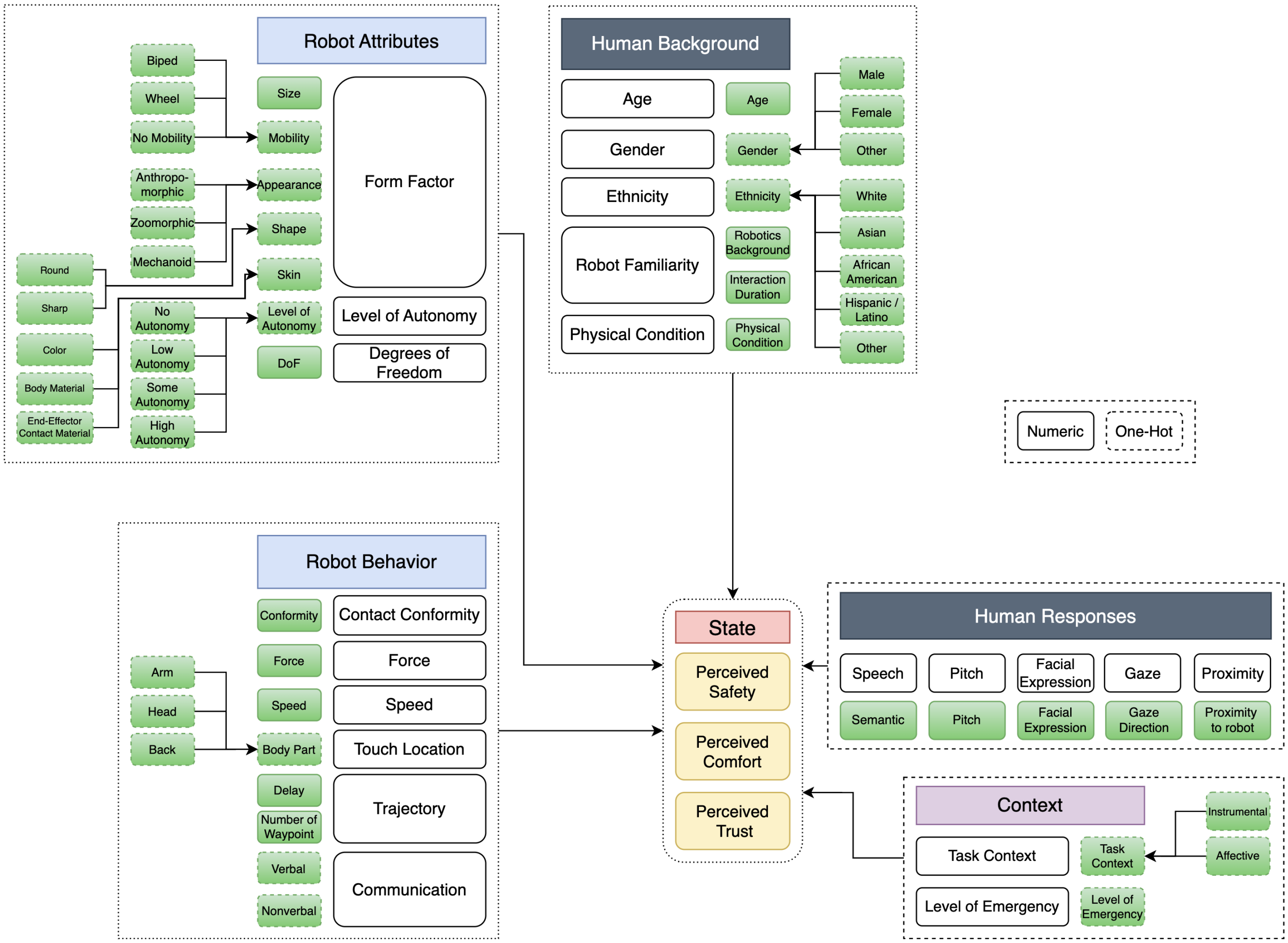

Figure 2: Perception Model for human state estimation based on human, robot, and context information.

Figure 3: Wiping arm and grooming hair tasks.

$$
\begin{aligned}
\max_{x_i \in X} \quad & \sum_{i=1}^{N} (\alpha b_i - \beta c_i) x_i \\
\text{s.t.} \quad & \sum_{i=1}^{N} t_i x_i \leq T \\
& x_i \in \{0, 1\} \\
& MutualExclusivity\ M
\end{aligned}
\tag{1}
$$

In our implementation, the benefit b_i of each behavior is represented with the ratings of safety, comfort, and intuitiveness of the behavior as perceived by the users. The cost c_i of each behavior is the ratings of the level of time-consumption and interruptiveness of the behavior as perceived by the users. The coefficients α and β are the dimensionless, relative scaling factors between the benefit and cost ($0 \leq \alpha, \beta \leq 1$). In our implementation, we set $\alpha = 1$ and $\beta = 0.5$ because we cared more about the behavior benefit than the cost. However, they can be customized in other specific applications and purposes. We use a *Benefit & Cost Predictor* which determines the benefits and costs of all the behaviors based on human background, context (task and level of emergency of the situation), and the current human state determined from the Perception Model. The pipeline is shown in Fig. 4. We used Dynamic Programming to solve the 0-1 Knapsack problem for its fast computation and guarantee of optimal solutions. Additionally, we consider $T = 10$ seconds in our specific implementation.

Data Collection and Model Evaluation

Perception Model

In a separate work, we have conducted a co-design experiment with 20 participants to understand users' perception of robot-initiated touch interactions based on varied robot parameters (speed, timing, smoothness of motion, number of contact points, force applied by the robot at contact, and contact material), and through two tasks- namely wiping arm to represent a medical scenario of sterilizing the area,

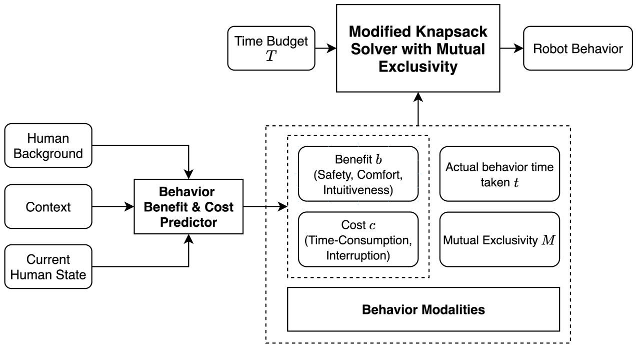

Figure 4: Behavior Adaptation pipeline, containing *Behavior Benefit & Cost Predictor* and *Knapsack Solver*. *Behavior Benefit & Cost Predictor* module provides the predicted benefit and cost of the available behaviors, and *Knapsack Solver* outputs the optimal behaviors to improve the interaction.

and grooming hair to represent a scenario of assistance with daily activities (Fig. 3)- with two trials per task. Order of tasks each participant would receive was randomized. The arm wiping task involved the robot using its hand to simultaneously apply pressure and sweep across participant's forearm. The hair grooming task involved the robot using its hand to stroke over participant's hair near forehead. Participants include 14 male and 6 female with age range (23-58), from different ethnicity, various levels of familiarity and experience working with robots (from having no exposure to robots to being a robotics researcher with daily exposures to running robots), and different levels of physical conditions (from some limitations in performing physical activities to no limitations). Each participant started with a practice task of *pressing wound* with robot hand to get familiar with the robot and different parameter choices. Then, for experiment tasks of wiping arm and grooming hair, for each task, in the first trial participants were asked to specify the values for each of the modifiable parameters of the robot, mentioned above, then the robot would execute the designed touch interaction for the participant to experience and evaluate the interaction in terms of *Perceived safety*, *Perceived Comfort*, and *Perceived Trust*, and report it through the survey and post-experiment interview. In the second trial, they were asked to modify as many of the parameters from first trial as they want in a way that they think will improve their perception. Then, the robot would execute the modified behavior and similar to first trial, the participant would evaluate and report on how they perceived the modified touch interaction. Thus, each trial of each participant was completely different from others, which highlights the variability and diversity of the collected dataset. All the related survey questions were based on 11-point Likert scale, ranging from 0 to 10. Noting that the main focus of the study was to explore what parameters participants selected to change and the resultant effect on their perception of interaction, and considering the already long 90-minute experiment session per participant for 4 trails, we believe two iterations for each task was reasonable. Moreover, due to the nature of a co-design study where participants could freely choose the robot parameters, we suspect that further evaluation (e.g., increased number of iteration per task to allow further rounds of change) would make major difference in data compared to the 2nd iteration, since people get more familiar with the robot and the task.

The hardware we used was a Franka Research 3 arm equipped with a QB SoftHand2. We used the data collected through this study - which contains *human background*, *human responses*, *robot behaviors*, *context* and the associated *human state ratings*- for training and evaluation of the Perception model. Since a single platform was used to conduct this user study, most of the *Robot Attributes* were fixed and except for "End-Effector Contact Material" they were not used for model training and evaluation. The study was approved by our institution's Institutional Review Board.

We used the data collected for the above experiments to train and evaluate a reduced version of the perception model-based on the subset of the parameters we had access to:

Robot Behavior are robot parameters, with "Contact Conformity" (discretized into 5 levels) representing contact surface between robot hand and body part, "Force", "Speed", "Delay", and "Number of Waypoint" are discretized into 5 levels according to their associated allowable range, "Body Part" is one-hot encoded for arm, head and back, and "Verbal" and "Nonverbal" are binary of whether the robot performs verbal and/or nonverbal behaviors.

Human Background were obtained from pre-experiment survey, in which "Age" is discretized into levels from 1 to 5, "Gender" and "Ethnicity" are one-hot encoded, "Robotics Background" is binary, "Interaction Duration" is the trial number (1 and 2), and "Physical Condition" is an integer from 1 (Have limited ability for physical activities) to 5 (Could do most physical activities effortlessly).

Human Responses were obtained from the interaction videos, in which "Semantic" and "Pitch" are binary depending on whether participant talks and has pitch change, "Facial Expression" is binary depending on whether participant shows any recognizable facial expression (happy, sad, etc), "Gaze" is binary depending on whether the human looks at the robot end effector, "Proximity" is integer from 1 to 3 representing the distance participant stands from the robot.

Context were obtained from the task description. "Task Context" is one-hot encoded between "Instrumental" and "Affective", and "Level of Emergency" (discretized into 5 levels from 1 to 5) is chosen as"1" (no emergency) for all the participants due to experimenting in a laboratory setting.

Human State were obtained from post-interaction survey, in which for each of *Perceived Safety*, *Perceived Comfort*, and *Perceived Trust*, participants rated from 0 to 10.

We explored different methods for regression, and ended up using Random Forest (Breiman 2001) to predict the continuous values of the human states on the evaluation set due to its better performance compared to the other two. The dataset contains 80 data points (20 participants, each undergoes 2 tasks with 2 trials each). We performed the 80/20 random split for collected data for training/evaluation purposes such that data of each participant (containing the entire 4 trials) were treated as one; thus we had whole data of 20% of participants (i.e. 4 participants) randomly selected for eval-

Metric	Perceived Safety	Perceived Comfort	Perceived Trust
MAE	0.842	0.779	0.933
MSE	1.177	0.820	1.086

Table 1: Perception Model Evaluation Results

uation, and the whole data of the other 16 participants for training. We did this to avoid any data leakage, where information from the same participant might appear in both the training and evaluation sets, to ensure generalizability. Moreover, for training, we used 5-fold cross validation to mitigate the effects of our rather limited sample size, in order to provide a more robust assessment of the model's performance and help prevent overfitting. The evaluation with Mean Absolute Error and Mean Squared Error is shown in Table 1. Noting that the safety, comfort, and trust ratings indicated by participants in our collected data ranged from 4 to 10, 3 to 10, and 3 to 10, respectively, the prediction errors are low, as shown in Table 1.

While we were able to evaluate a reduced version of the proposed perception model through the collected data (without most of *Robot Attributes*), the results of current evaluation indicate that the perception model has the capability to estimate human states with proposed parameters.

Communication Behavior Adaptation

For evaluating feasibility of the behavior adaptation module, we conducted a user study to collect some data on users' preference for robot's nonverbal and verbal communication behaviors while physically interacting with humans. The behaviors are described in Table 2. These communication behaviors are adapted from HRI literature on robot behaviors (Ngo, Carter, and Steinfeld 2024; Han, Phillips, and Yanco 2021; Rawal and Stock-Homburg 2022; Kshirsagar et al. 2020) as well as desired behaviors extracted from the post-experiment interviews of the user study in . Of the 20 participants who took part in the study described in Perception Model section, 16 (11 male, 5 female and age range (27-58)) participated in this data collection.

Considering the two touch tasks in IV.A, we asked the participants to rate how they would perceive the interaction if the robot showed any of the communication behaviors in Table 2 and during different timings of interaction such as ”Before Touch”, ”During Touch”, and ”Before First Contact”. We considered “Eye gaze at user” and “Eye gaze at touch location” as two separate behaviors. Thus, the actual total number of behaviors in our dataset was 20. We collected 1600 data in total, consisting of 16 participants, 20 behaviors, and 5 human perception of the behaviors (safety, comfort, intuitiveness, time-consumption and interruptiveness). The questions targeted at adding specific behaviors and asked whether each behavior in question would contribute to the participant's feeling of safety, comfort, and intuitiveness in the interaction with the robot and how time-consuming and interruptive they find that behavior to be. The questionnaire was designed using the Godspeed Questionnaire (Bartneck et al. 2009) and the robotic social attributes scale (Carpinella et al. 2017). All questions were based on 7-point Likert scale ranging from 1 (Strongly Disagree) to 7 (Strongly Agree). The study was approved by our institution's Institutional Review Board.

Group	Behavior	Description
Verbal	Greetings	The robot verbally greets users
	Prescreen	Robot asks user some background questions
	Permission	Robot asks user for permission before touching
	Intention Description	Robot describes the intention of the touch to user before touching
	Notification	Robot notifies user before ending the touch
	Feedback	Robot asks user for feedback on the touch
Nonverbal	Eye Gaze	Robot *looks at user* or *looks at touch location*
	Facial Expression	Robot shows facial expression, either *static* or *dynamic*, during the touch
	Hand Showing	Robot shows its hands to user (*before first contact, before each contact, after touch,* or *before changing direction*)
	Short Pause	Robot pauses shortly right before the contact to obtain unsaid permission (*before first contact, before each contact,* or *before finishing touch*)
	Light Tap	Robot taps lightly on user to signal intention (*before first contact, before each contact,* or *before finishing touch*)

Table 2: Robot behaviors and descriptions

Using the collected data, we trained a ”Behavior Benefit & Cost Prediction” model which predicts the benefits (average value of *Perceived Safety*, *Comfort*, and *Intuitiveness*) and costs (average value of *Time-consumption* and *Interruptiveness*) associated with each behavior, given the information on human's current state, human background, and context (“task” and “level of emergency”). The output is a pair of real numbers from 1 to 7. For the human's state and background, and task context, we used the data collected in IV.A associated with each participant and task, and the human state values are associated estimated values provided by the output of the Perception Model.

Similar to the Perception Model, the training set for the ”Behavior Benefit & Cost Prediction” model consisted of 80% of data where data was split randomly and was participant-specific: whole data of randomly selected 13 participants were used for training and whole data of remaining 3 participants were used for evaluation. We used both deep-learning (TabTransformer (Huang et al. 2020), TabNet (Arik and Pfister 2021)) and tree-based methods (Random Forest (Breiman 2001), XGBoost (Chen and Guestrin 2016)) for our model training and evaluation (fine-tuned with Random Search (Bergstra and Bengio 2012)). For comparison metrics, we used the average MAE and MSE values for all robot behaviors on our evaluation set (data from 3 participants). Specifically, we calculated the MAE and MSE of each behavior (with associated interaction timing if available). Then, we took the average MAE and MSE values on all 20 behaviors for 3 evaluation participants. Thus, the final MAE and MSE values for behavior benefit and cost prediction were represented per behavior per participant. Among the methods, XGBoost performed the best overall, with benefit prediction MAE of 0.743 and MSE of 0.959, and cost prediction MAE of 0.945 and MSE of 1.468.

Finally, using the predicted benefits and costs of behaviors, along with time budget ($T = 10$) and Mutual Exclusiv-

Rank	Perceived Safety		Perceived Comfort		Perceived Trust	
	Feature	Value	Feature	Value	Feature	Value
1	Proximity	0.0913	Age	0.1801	Interaction Duration	0.0842
2	Body Part	0.0826	Physical Condition	0.1299	Contact Conformity	0.0831
3	Contact Conformity	0.0762	Body Part	0.0906	Facial Expression	0.0694
4	Interaction Duration	0.0238	Proximity	0.0706	Proximity	0.0574
5	Age	0.0235	Interaction Duration	0.0477	Age	0.0519

Table 3: Human State Contributors (Mean Absolute SHAP)

ity M, through the modified Knapsack Solver (described in Communication Behavior Adaptation section) the *behavior adaptation* module outputs the optimal robot communication behaviors to improve human perception. For example, the model output of one of our participants (White Male, has robotics background and rated "5" on physical condition, age from 45 to 60) was "Greetings", "Permission", "Intention Description", "Gaze at touch location", "Dynamic Facial Expression", "Hand showing first contact", and "Light Tap first contact". Their outputs consist of 3 out of 6 verbal behaviors and 4 out of 14 nonverbal behaviors, which indicate their preference towards verbal behaviors and consistent with our previous analysis.

Design Insights for pHRI

We performed SHAP Feature Importance analysis (Lundberg 2017) to quantitatively determine which features contribute the most to the human state estimation, based on feature's mean SHAP values on all data points averaging across *Perceived Safety*, *Perceived Comfort*, and *Perceived Trust*. The top 5 most important features are shown in Table 3.

Results show that "Proximity", "Robotics Background", and "Age" are among the most important features determining human state. This indicates that people with different robotics background and age could have different perceptions of robots, and their proximity to the robot is a good indicator of their state during interaction. In addition to those parameters, "Body Part" (touch location) and "Contact Conformity" (how much the robot hand conforms to the human body part) also have high influence on human perception. Such insights of "Proximity" and "Body Part" are consistent with previous works (Zheng et al. 2019; Rubagotti et al. 2022), validating feasibility of our proposed perception model. In addition to the mean absolute SHAP values, some directional contributions of the features to the output can also be concluded from mean signed SHAP values (i.e., a positive mean SHAP value means increasing that feature also increases the output, and vice versa). For example, higher "Proximity" resulted in higher perceived safety, indicating that participants standing further from the robot felt safer during interaction. Participants having higher "Physical Condition" felt more comfortable interacting with the robot. Higher SHAP values for "Contact Conformity" and "Physical Condition" also verify our hypothesis that these proposed parameters to the perception model by us (non-literature based) are valid to be considered in the perception model. Higher "Interaction Duration" increases trust, as participants trust the robot more if they spend more time with it, which is consistent with literature where human perception such as trust is a function over interaction time (Lewis, Sycara, and Walker 2018). Moreover, having visible facial expressions and lower contact conformity (fingers wrap less around contact area) indicate a lower level of trust. Moreover, with "Age", older participants had higher perceived safety and comfort level, but lower trust, which make sense as people of different ages form their perceptions towards robots differently (Morillo-Mendez et al. 2022). Finally, facial expression is a good indicator of human perception towards robot, as participants might show some recognizable facial action when touched by robots, such as eyebrow raising (Hoffmann and Krämer 2021). All the above insights suggest that researchers should consider task related parameters (i.e., "Contact Conformity", "Body Part", "Proximity", etc) to inform the design of robotic touch behaviors. Factors such as "Level of Emergency" or "Task Context" did not show up in the top of important features, due to all participants being under the same condition and situation. However, the importance of those factors are supported by literature (Rossi et al. 2017; Chen et al. 2011), thus their inclusion in our perception model is still valid.

In addition to SHAP analysis, we performed ablation study to see if excluding any feature could worsen the safety, comfort, or trust prediction results. The result is shown in Table 4. We observed that including human background (i.e., "Age", "Physical Condition") improves the prediction results of safety and comfort. Moreover, robot behavior (i.e., "Speed", "Delay", "Number of Waypoint") is needed in predicting safety, and "Body Part" is valuable in both safety and comfort prediction. Finally, neglecting human responses (i.e., "Speech", "Pitch", "Facial Expression", and "Gaze") worsen the results of safety prediction. Thus, part of robot attributes (e.g., "End-Effector Contact Material"), robot behavior, human background, and human responses are important for more accurate human state prediction.

Finally, we performed two t-tests to compare benefit and cost in verbal behaviors and nonverbal behaviors. Nonverbal behaviors ($M = 3.952, SD = 0.596$) have higher cost than verbal behaviors ($M = 3.224, SD = 1.031$); $t(15) = -2.368, p = .025$. Such result suggests that verbal behaviors are a more preferred and efficient way of communication, which is consistent with literature (Ngo, Carter, and Steinfeld 2024; Han, Phillips, and Yanco 2021).

Discussion

Perception Model

Although previous works have investigated how human state is influenced by human background, robot parameters, and interaction context (Akalin, Kristoffersson, and Loutfi 2022; Hoffmann and Krämer 2021; Hu et al. 2022; Kanda et al. 2008), there hasn't been an effort on building a consolidated multi-modal perception model for the purpose of determining human states, especially during robot-initiated touch scenarios. We presented the consolidated knowledge from the literature through a multi-modal model that can be applied to most robot-initiated touch scenarios. While full

Model Variant	Safety MAE	Safety MSE	Comfort MAE	Comfort MSE	Trust MAE	Trust MSE
All Features	0.8420	1.1772	0.7792	0.8200	0.9333	1.0863
Skip “Age”	0.8473	1.1854	0.8317	0.9226	0.8947	1.0259
Skip “Gender”	0.8461	1.1768	0.7756	0.8603	0.8999	1.0698
Skip “Ethnicity”	0.7924	1.0855	0.7499	0.7278	0.9020	0.9894
Skip “Robotics Background”	0.8329	1.1663	0.8043	0.8691	0.8827	1.0536
Skip “Interaction Duration”	0.8351	1.1816	0.7269	0.7309	0.7440	0.7912
Skip “Physical Condition”	0.8493	1.2027	0.8335	0.9054	0.8313	0.9119
Skip “End-Effector Contact Material”	0.8454	1.1814	0.7309	0.7170	0.8204	0.8810
Skip “Contact Conformity”	0.8844	1.2286	0.7585	0.7585	0.8441	0.9289
Skip “Force”	0.8274	1.1425	0.7377	0.7052	0.8641	0.9346
Skip “Speed”	0.8529	1.1872	0.7787	0.7723	0.8390	0.9116
Skip “Body Part”	0.9072	1.2464	0.7844	0.8265	0.8902	0.9993
Skip “Delay”	0.8514	1.1797	0.7780	0.8281	0.8857	0.9898
Skip “Number of Waypoint”	0.8636	1.2092	0.7326	0.7302	0.8433	0.8914
Skip “Verbal”	0.8584	1.2089	0.7836	0.8528	0.8685	0.9519
Skip “Nonverbal”	0.8584	1.2089	0.7836	0.8528	0.8685	0.9519
Skip “Speech”	0.8669	1.2047	0.7876	0.8595	0.8810	1.0093
Skip “Pitch”	0.8783	1.2563	0.7434	0.7539	0.8611	0.9422
Skip “Facial Expression”	0.8680	1.2448	0.7258	0.7292	0.8631	0.9422
Skip “Gaze”	0.8856	1.2718	0.7519	0.7526	0.8880	1.0205
Skip “Proximity”	0.7570	0.9782	0.7326	0.7302	0.8671	0.9446
Skip “Task Context”	0.8562	1.1952	0.7744	0.7864	0.9022	1.0096
Skip “Level of Emergency”	0.8697	1.2138	0.7316	0.7145	0.8737	1.0026

Table 4: Ablation Study Results for Perception Model

evaluation of the proposed model under various scenarios and parameters is a research work of its own- which we plan to do in the future- nevertheless we were able to show the feasibility of such a model by evaluating a reduced size model through user study data. The low estimation errors of the perception model along with feature importance analysis reinforced the potential of our proposed model on providing human state estimates. Future researchers can use this model to adapt to their specific applications, as well as design a more comprehensive data collection for their experiments.

In our evaluation, we did not consider *Robot Attributes*, except for “End-Effector Contact Material”, due to having access to a single robot platform. Thus, we couldn’t quantitatively validate most of the robot attributes portion of the perception model. Nevertheless, literature supports robot attributes effect on human perception (Gutman, Olatunji, and Edan 2021; Rubagotti et al. 2022; Kanda et al. 2008). We ground our hypothesis on existing literature that robot attributes are contributing factors in human state, and call for future work of evaluation across robot platforms to further evaluate the proposed model with varying robot attributes.

Moreover, due to the nature of a co-design study and time limitations for experiments to minimize participant fatigue, the perception model dataset contain 2 trials per task per participant. While we believed that such number of evaluations per participant was informative enough for the current purpose of the study, future work could consider expanding the experimental design to collect a broader range of robot motion patterns to provide an even more accurate assessment on how different robot behaviors affect human perception. In an extended effort, longitudinal data could also be collected to assess the long-term effectiveness of such proposed parameters in the Perception Model.

Finally, although exclusion of physiological measurements in Perception Model section expands the usable scenarios of this model due to lack of dependency on such sensors, physiological data may further enhance the effectiveness of the Perception Model. Such evaluations need to be studied in the future by using data from wearable sensors.

Behavior Adaptation

Despite the inclusion of adaptation for robot parameters in the system overview section, we only implemented the communication behavior adaptation side in this paper. However, a similar procedure to that of communication adaptation can be followed for the robot parameter adaptation, in which participant’s preference on changes of the robot parameters with respect to a nominal value can be captured through data collection to train a benefit & cost prediction model. The Knapsack problem formulation and behavior benefits and costs prediction can still be applied.

Finally, human’s perception of interaction after the recommended behaviors are implemented requires further user studies to evaluate whether they actually improved the user’s perception or not. This will be studied in our future work.

Conclusion

We introduced a multi-modal perception model for human state estimation in physical human-robot interaction that can be applied and adapted to many different contexts, including the less explored robot-initiated touch. We proposed a framework for verbal and nonverbal behavior adaptation for recommending behaviors to improve human perception during interaction while considering costs and constraints. While extensive evaluation of the models should be performed in future studies, the introduced models and the insights gained through the feature analysis, should guide the design of physical human-robot interaction systems.

References

Akalin, N.; Kristoffersson, A.; and Loutfi, A. 2022. Do you feel safe with your robot? Factors influencing perceived safety in human-robot interaction based on subjective and objective measures. *International journal of human-computer studies*, 158: 102744.

Alonso-Martin, F.; Malfaz, M.; Sequeira, J.; Gorostiza, J. F.; and Salichs, M. A. 2013. A multimodal emotion detection system during human–robot interaction. *Sensors*, 13(11): 15549–15581.

Arik, S. Ö.; and Pfister, T. 2021. Tabnet: Attentive interpretable tabular learning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, 6679–6687.

Bartneck, C.; Kulić, D.; Croft, E.; and Zoghbi, S. 2009. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International journal of social robotics*, 1: 71–81.

Beer, J. M.; Fisk, A. D.; and Rogers, W. A. 2014. Toward a framework for levels of robot autonomy in human-robot interaction. *Journal of human-robot interaction*, 3(2): 74.

Bergstra, J.; and Bengio, Y. 2012. Random search for hyper-parameter optimization. *Journal of machine learning research*, 13.

Breiman, L. 2001. Random forests. *Machine learning*, 5–32.

Carpinella, C. M.; Wyman, A. B.; Perez, M. A.; and Stroessner, S. J. 2017. The robotic social attributes scale (RoSAS) development and validation. In *Proceedings of the 2017 ACM/IEEE International Conference on human-robot interaction*, 254–262.

Chen, T.; and Guestrin, C. 2016. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 785–794.

Chen, T. L.; King, C.-H.; Thomaz, A. L.; and Kemp, C. C. 2011. Touched by a robot: An investigation of subjective responses to robot-initiated touch. In *Proceedings of the 6th international conference on Human-robot interaction*, 457–464.

Chen, T. L.; King, C.-H. A.; Thomaz, A. L.; and Kemp, C. C. 2014. An investigation of responses to robot-initiated touch in a nursing context. *International Journal of Social Robotics*, 6: 141–161.

Fréville, A. 2004. The multidimensional 0–1 knapsack problem: An overview. *European Journal of Operational Research*, 155(1): 1–21.

Gouaillier, D.; Hugel, V.; Blazevic, P.; Kilner, C.; Monceaux, J.; Lafourcade, P.; Marnier, B.; Serre, J.; and Maisonnier, B. 2009. Mechatronic design of NAO humanoid. In *2009 IEEE international conference on robotics and automation*, 769–774. IEEE.

Gutman, D.; Olatunji, S.; and Edan, Y. 2021. Evaluating levels of automation in human–robot collaboration at different workload levels. *Applied Sciences*, 11(16): 7340.

Han, Z.; Phillips, E.; and Yanco, H. A. 2021. The need for verbal robot explanations and how people like a robot to explain itself. *ACM Transactions on Human-Robot Interaction*, 10: 1–42.

Hirano, T.; Shiomi, M.; Iio, T.; Kimoto, M.; Tanev, I.; Shimohara, K.; and Hagita, N. 2018. How do communication cues change impressions of human–robot touch interaction? *International Journal of Social Robotics*, 10: 21–31.

Hoffmann, L.; and Krämer, N. C. 2021. The persuasive power of robot touch. Behavioral and evaluative consequences of non-functional touch from a robot. *Plos one*, 16(5): e0249554.

Hu, Y.; Abe, N.; Benallegue, M.; Yamanobe, N.; Venture, G.; and Yoshida, E. 2022. Toward active physical human–robot interaction: Quantifying the human state during interactions. *IEEE Transactions on Human-Machine Systems*, 52(3): 367–378.

Hu, Y.; Benallegue, M.; Venture, G.; and Yoshida, E. 2020. Interact with me: An exploratory study on interaction factors for active physical human-robot interaction. *IEEE Robotics and Automation Letters*, 5(4): 6764–6771.

Huang, X.; Khetan, A.; Cvitkovic, M.; and Karnin, Z. 2020. Tabtransformer: Tabular data modeling using contextual embeddings. *arXiv preprint arXiv:2012.06678*.

Kanda, T.; Miyashita, T.; Osada, T.; Haikawa, Y.; and Ishiguro, H. 2008. Analysis of humanoid appearances in human–robot interaction. *IEEE transactions on robotics*, 24(3): 725–735.

King, C.-H.; Chen, T. L.; Jain, A.; and Kemp, C. C. 2010. Towards an assistive robot that autonomously performs bed baths for patient hygiene. In *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 319–324. IEEE.

Kshirsagar, A.; Lim, M.; Christian, S.; and Hoffman, G. 2020. Robot gaze behaviors in human-to-robot handovers. *IEEE Robotics and Automation Letters*, 5(4): 6552–6558.

Lewis, M.; Sycara, K.; and Walker, P. 2018. The role of trust in human-robot interaction. *Foundations of trusted autonomy*, 135–159.

Lundberg, S. 2017. A unified approach to interpreting model predictions. *arXiv preprint arXiv:1705.07874*.

Magyar, G.; and Vircikova, M. 2015. Proposal of a cloud-based agent for social human-robot interaction that learns from the human experimenters. In *2015 IEEE 19th International Conference on Intelligent Engineering Systems (INES)*, 107–111. IEEE.

Martello, S.; and Toth, P. 1987. Algorithms for knapsack problems. *North-Holland Mathematics Studies*, 132: 213–257.

Mazursky, A.; DeVoe, M.; and Sebo, S. 2022. Physical touch from a robot caregiver: Examining factors that shape patient experience. In *2022 31st IEEE International conference on robot and human interactive communication (RO-MAN)*, 1578–1585. IEEE.

Mitsunaga, N.; Smith, C.; Kanda, T.; Ishiguro, H.; and Hagita, N. 2008. Adapting robot behavior for human–robot interaction. *IEEE Transactions on Robotics*, 24(4): 911–916.

Morillo-Mendez, L.; Schrooten, M. G.; Loutfi, A.; and Mozos, O. M. 2022. Age-related differences in the perception of robotic referential gaze in human-robot interaction. *International journal of social robotics*, 1–13.

Ngo, H. Q.; Carter, E. J.; and Steinfeld, A. 2024. Human Perception of Robot Failure and Explanation During a Pick-and-Place Task. In *Proceedings of the AAAI Symposium Series*, volume 4, 373–379.

Ngo, H. Q.; and Steinfeld, A. 2024. Joint Potential-Vector Fields for Obstacle-Aware Legible Motion Planning. In *2024 33rd IEEE International Conference on Robot and Human Interactive Communication (ROMAN)*, 1856–1863. IEEE.

Ozsoydan, F. B.; and Gölcük, İ. 2023. A reinforcement learning based computational intelligence approach for binary optimization problems: The case of the set-union knapsack problem. *Engineering Applications of Artificial Intelligence*, 118: 105688.

Peña, D.; and Tanaka, F. 2020. Human perception of social robot's emotional states via facial and thermal expressions. *ACM Transactions on Human-Robot Interaction (THRI)*, 9(4): 1–19.

Rawal, N.; and Stock-Homburg, R. M. 2022. Facial emotion expressions in human–robot interaction: A survey. *International Journal of Social Robotics*, 14(7): 1583–1604.

Rossi, A.; Dautenhahn, K.; Koay, K. L.; and Walters, M. L. 2017. How the timing and magnitude of robot errors influence peoples' trust of robots in an emergency scenario. In *Social Robotics: 9th International Conference, ICSR 2017, Tsukuba, Japan, November 22-24, 2017, Proceedings 9*, 42–52. Springer.

Rubagotti, M.; Tusseyeva, I.; Baltabayeva, S.; Summers, D.; and Sandygulova, A. 2022. Perceived safety in physical human–robot interaction—A survey. *Robotics and Autonomous Systems*, 151: 104047.

Shiomi, M.; Hirano, T.; Kimoto, M.; Iio, T.; and Shimohara, K. 2020. Gaze-height and speech-timing effects on feeling robot-initiated touches. *Journal of Robotics and Mechatronics*, 32(1): 68–75.

Thongsri, N.; Warintarawej, P.; Chotkaew, S.; and Saetang, W. 2022. Implementation of a personalized food recommendation system based on collaborative filtering and knapsack method. *Int. J. Electr. Comput. Eng*, 12(1): 630–638.

Yutaka, M.; and Ken, T. 2005. Emotion detecting method based on various attributes of human voice. *Science of Emotion and Sensibility*, 8(1): 1–7.

Zheng, X.; Shiomi, M.; Minato, T.; and Ishiguro, H. 2019. What kinds of robot's touch will match expressed emotions? *IEEE Robotics and Automation Letters*, 5(1): 127–134.