Human-Robot Interaction: A State of the Art

Souhila KAHLOUCHE
Centre de développement des Technologies Avancées (CDTA)
Cité 20 août 1956- Baba Hassen, Alger, Algérie
skahlouche@cdta.dz

Abstract: The introduction of robots into our daily lives has created new challenges for autonomous robots: adapting to the existence of humans in the same environment and interacting with them has given rise to a new research field called Human-Robot Interaction (HRI), where the robot is considered as a partner rather than a tool. This survey provides a structured overview of current research on HRI. A detailed taxonomy of this research axes is presented, highlighting contemporary work in verbal and non-verbal interaction, emotional expression, intention recognition, social navigation, and human perception. For each axis, we discuss seminal works, prevailing methodologies, and open challenges.

Keywords: Human-Robot Interaction (HRI); Social Robotics; Service Robots; Robot Autonomy; Verbal Interaction; Non-Verbal Interaction; Social Navigation; Intention Recognition; Human Activity Recognition.

1. INTRODUCTION

Usually, autonomous mobile robots have been deployed in areas where interaction with humans is not fundamental, such as the automotive industry, oil well inspection, and exploration of environments hostile to humans. These robots are generally supervised by a human operator.

Recently, service robots are increasingly present in our environment and evolving rapidly to acquire more and more important roles where social interaction is a significant part of the robot's activity.

Examples include guide robots in airports or large commercial centers, receptionist robots in public places, or hospital worker robots or caregiver robots for the elderly [1].

This leads to the emergence of the Human–Robot Interaction (HRI) research field that aims to enable robots to perform complex tasks in close collaboration with humans.

Figure 1 shows some service robot platforms:

- Starship Technologies: Autonomous, sidewalk robots that deliver food and parcels on university campuses and in city districts.
- Bear Robotics: Penny the robot waiter, designed to help servers.
- Amazon Robotics (Proteus, Sequoia, etc.): A vast fleet of autonomous mobile robots (AMRs) that move inventory, sort packages, and power Amazon's fulfillment centers.



Figure 1: Examples of service robots.

It is necessary to identify some important aspects when designing an HRI system. These aspects are:

- *Proximate or remote communication,* according to if the human and robot are close or far apart.
- *The Human/Robot team structure,* it is required to define the human role in the interaction as *Supervisor, Partner, etc.* to identify when the human must control the entire system and when the human and the robot work together to accomplish a common objective [2].
- *Nature of Exchanged Information* must be determined in HRI system, how information is exchanged between human and robot? [3] means (sensors, network) and format (Visual, Speech, haptic) [4] .
- *Level of Robot Autonomy,* it describes to what degree the robot can act on its own. This level varies from the robot completely controlled by the human (tele-operated) to the completely autonomous robot not needing to have its actions approved by the human.

In essence, the paper's primary contribution is to present a detailed taxonomy of active research axes in HRI field, covering verbal/non-verbal interaction, emotion, intention, navigation, and perception, and also to identify

convergent trends and promising directions for future HRI research. Therefore, this paper will serve as a reference for researchers entering this field.

2. TAXONOMY OF RESEARCH WORKS IN HRI

Social interaction between a human and a robot is characterized by a continuous exchange of gestural, emotional, and communicative information.

Many researchers believe that human interaction with their peers when face-to-face frequently goes through non-verbal communication.

Equipping a conventional robot, capable of perceiving its environment, planning and executing tasks, with human interaction capacity, does mean giving it the ability to understand, in addition to natural language, the tools to perceive the emotions and social signals we emit, to understand and predict human intentions, and to adapt to change.

According to a synthesis published by Thomaz [5]. Research works in HRI is classified under several sub-axes of research. Figure 2 shows a non-exhaustive list of these sub-axes in the HRI domain.

A. Verbal Interaction: Speech is one of the most studied means of communication, as it allows the transfer of information rich in content between the robot and its interlocutor. Nevertheless, this communication faces many problems related to parameters and interaction conditions..

Some parameters are related to speaker dependence, such as familiar expressions, varied pronunciations of the same word, vocabulary size, diversity of expressions used by different speakers, etc. Other problems are related to the interaction environment such as noise, incompatibility of conditions during the learning and use phase, as well as discrete or continuous recognition, etc

The major difference between human-robot dialogue, to produce correct vocal content in an HRI scenario, and general dialogue of AI systems, is that it depends on the physical environment, the collaboration situation detected by a variety of sensors, as well as human spatial behavior. Because a social robot must understand language in the physical context in which it is located.

There is also the interaction system for guide robots and city exploration [6] where the robot uses gesture and touch screen as input modalities and uses speech and image as output modalities. In the majority of these works, the robot generates speech from a text string, which is a relatively easy thanks to speech synthesis engines. However, recognizing speech is much more difficult, as it involves additional uncertainty on the input (noise, accents, timing) beyond semantic content. For this reason, many robots speak but do not understand spoken language.

This asymmetry between produced (generated) language and perceived language constitutes a problem in the social relationship between humans and robots, because people expect from a social robot, capable of generating relatively complex speech, that it can understand speech of similar complexity.

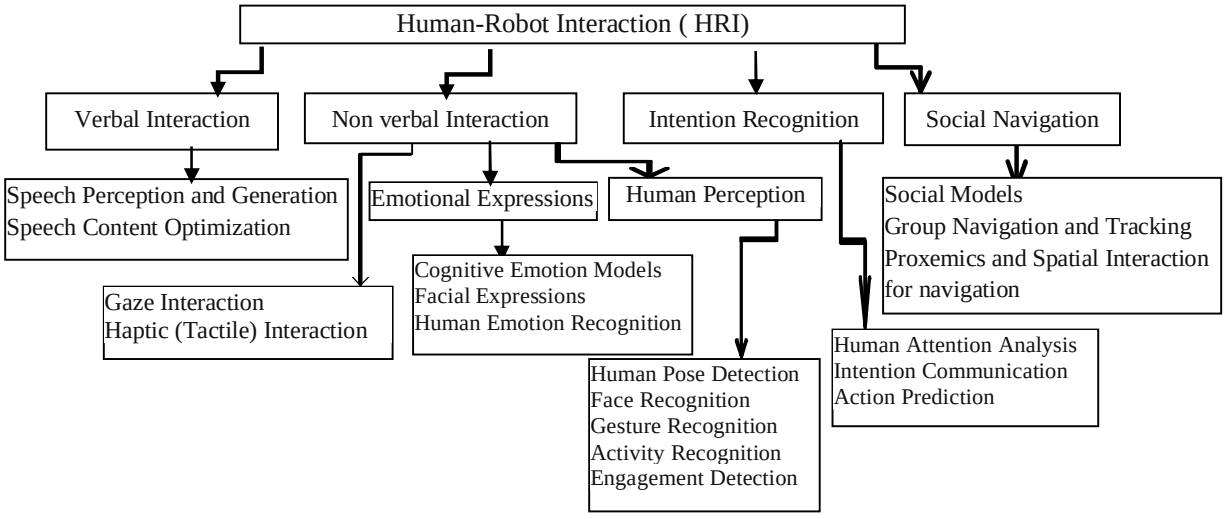

Figure 2: Main research works in HRI

In the framework of collaborative human-robot tasks, in [7], the authors created a system based on an information-theoretic strategy to generate clarification questions to remove ambiguity. Furthermore, robots ask for help when a failure is detected during an assembly task.

The help request is optimized to facilitate understanding by a human listener using "inverse semantics," which employs a probabilistic model of the human process of interpreting a request.

Paralinguistics is another aspect highly studied by researchers in verbal interaction with social robots. It concerns properties of verbal behavior that are integrated into the spoken message [8] based on research on politeness and informal discourse in humans.

The authors in [9] propose ways to give advice offering the robot less authoritative behavior. This involves using expressions like "I think," "maybe," "somewhat" or discourse markers like "you know," "I mean," "well," "just," "like" in their instructions.

In summary, research in speech generation applied to HRI has focused on how to represent knowledge [10] , on speech content optimization algorithms [11] (particularly for explanations and referential expressions), and on algorithms to generate paralinguistic cues and timing parameters that determine how speech is delivered [12] .

On the other hand, research in speech understanding in a human-robot social dialogue addresses not only the semantic aspect of interaction but can also be helped by taking advantage of the nature of the interaction, which can help frame and contextualize the recognition problem.

B. Non-Verbal Interaction: To reinforce social interaction with humans, the robot must understand and control certain modalities related to non-verbal behavior [13]. Indeed, a large part of research in the literature of systems that produce and recognize non-verbal behavior deals mostly with recognizing pointing gestures and eye gaze.

To recognize human illustrative gestures, [14] proposed a spatial reasoning system taking into account eye-hand relationships with the environment to respond to the problem of pointing towards an object or direction from an RGB-D camera. In the same context, [15] propose a technique to consolidate recognition of pointing gestures and eye gaze where their study shows that people generally do not point to designate a person, but rather use gaze. A robot is considered more polite when it balances both types of pointing in a social situation

In order to automatically generate illustrative gestures, [16] use a learning approach to equip the robot with communicative behavior from pragmatic actions. Using a Hierarchical Markov Decision Process and a set of primitive behaviors including visual feature tracking and object search and manipulation, the robot learns that appropriate addition of eye gaze improves the performance of human pointing communication.

In another learning-based system, [17] use Gaussian Process regression to teach the robot pointing direction from a set of body features extracted from depth and amplitude images of a Time-of-Flight (ToF) camera.

Some works have proposed systems that combine verbal and non-verbal communication. [18] present a system allowing a museum robot to move its head to "interactionally significant" points during a presentation, such as transition points, in response to a question, on keywords, or with unknown words, in conjunction with illustration words. This combination of modalities demonstrated a positive effect on human engagement in the interaction.

Another non-verbal communication channel, as important as demonstrative gestures and eye gaze, is tactile detection and touch generation. Several robotic systems infer human intentions through force detection on the robot's end-effectors. For example, in [19] the robot "feels" the human pulling it and infers the vector and speed of the intended movement.

Researchers have also explored tactile sensors for robotic skins. An example is the "Huggable" robot skin [20], and the artificial fur developed by [21] (see figure 3). During their work, [22] design an extensible sensitive skin and a classifier based on the LogitBoost algorithm to classify six emotions and six social messages transmitted by humans touching an artificial arm.



Figure 3: Huggable Robot

Some works instead focus on the 'handshake' gesture. For example, [23] show that joint rigidity and grasp force can modulate the feeling of dominance emitted by a robot, and [24] find that shaking hands before negotiations improves cooperation.

In summary, non-verbal behavior is one of the most studied domains in HRI where a large part of research focuses on a subset of behaviors, notably on demonstrative or illustrative gestures, and eye gaze [25].
The subject of haptics in HRI [5], or interacting with robots through touch [26], is also suitable for future research, particularly in the context of robots touching humans (see figure 4), or in exploring new materials.
These aspects offer promising perspectives for future research on non-verbal behavior for HRI.

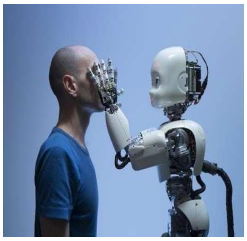

Figure 4: Robot with Haptic Interaction

C. Emotional Expression: Among the very important aspects of HRI, communication through expression and recognition of emotions stands out and plays a major role.
The relationship between emotions and HRI on one hand and technology on the other is highlighted by [27] who argues that emotional interaction should play a central role in the design of technological artifacts, particularly in robot design. A concrete example of an emotionally expressive social robot is the Kismet robot, which is an anthropomorphic robot with exaggerated emotion producing a set of facial expressions allowing it to express certain emotions [28] (see Figure 5). Kismet's attention system includes drivers and affective states, where the dynamic states of the drives affect the robot emotional state. This dynamic plays out in the robot's emotional behavior, resulting in continuous social exchange with the human partner.

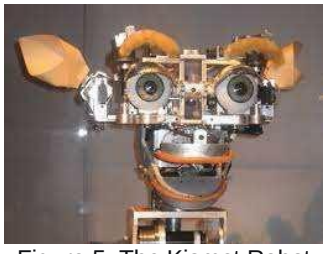

Figure 5: The Kismet Robot

Many researchers are influenced by and build upon Kismet's emotion architecture. [29] present a Framework that models five different functions of emotion, namely regulatory, selective, expressive, motivational, and ranking functions.

In the same way as emotional expression, there are works that address the inverse problem: detecting emotions in a social context. Indeed, recognizing the emotional state of a human during interaction with a robot has several uses. In the framework of collaboration, it can allow a robot to adapt its behavior to the state and preferences of its human partner. In a learning interaction, it can serve as a natural channel for human feedback on the execution of newly learned actions by the robot.
Several researchers have been interested in extracting facial expressions and facial features through vision during human-robot interaction. For example, [30] proceed to automatic recognition of facial communicative signals in the framework of an object teaching scenario.
In other works, the real-time emotion recognition system locates faces and extracts their features in an image sequence. It then codes facial expressions into one of seven different emotional states: happiness, sadness, fear, disgust, anger, surprise, and neutrality [31].
In the visual modality, emotions can be determined by characteristics other than facial expressions. [32] analyze video characteristics extracting postures and body movements to detect the engagement of children playing chess with a robot.
To summarize, in the HRI domain, research work related to emotions focuses on three axes: the design of emotional and affective behavioral models; the design of emotion expression and communication systems, mainly through facial expressions; and the design of systems to detect emotions in humans during interaction with a robot. However, these three axes (models, expression, and recognition) have for the most part been studied in isolation from each other and outside the context of whole interactions or collaboration tasks. This suggests that an approach combining the three aspects of emotions in HRI could be fertile ground for future research work.

D. Intention Recognition: This aspect aims to equip the robot with a high-level skill which is inferring human intentions and expressing its own intentions. Indeed, humans have a natural tendency to interpret others' behavior as intentional actions directed toward a goal. The ability to interpret another person's mental state is called theory of mind or "ToM," a capacity that exists in children from a very young age (e.g., following, joint attention, and social referencing), and develops progressively over the years [33].
Several cognitive architectures for HRI have been inspired by the notion of ToM and have

attempted to implement this aspect to allow robots to analyze intentional behavior to establish joint attention between human and robot [34].

For example, [35] describe a cognitive architecture for HRI allowing the robot to build a model to understand how and why people think as they do. The authors give three examples of specific social situations in which this architecture is capable of representing and reasoning notably: spatial perspective taking, action perspective taking in a team reasoning task, and dominant-submissive reasoning in a social task. In all three cases, the system uses its own competence model to understand a given situation.

In the context of intention analysis in a shared workspace between the robot and its human collaborator, [36] calculate "mightability maps" allowing a robot to reason about what its human collaborator could see or reach in the shared space. This map is calculated using the human's position, posture, and visual perspective taking.

ToM or the ability to infer a variety of mental states is a faculty that most social robots can achieve today [5]. It is clearly the precursor to a large part of social interactions; where the robot must infer, represent, and reason about aspects of the environment during an interaction.

Another challenge for the HRI research community for many years is attention analysis. [34] describe a tripartite model for joint attention capacity for social robots namely: responding to joint attention, initiating joint attention, and ensuring joint attention. In their study, they show that responding to joint attention improves performance in an object labeling task, and the robot is perceived as more competent and socially interactive. In a second experiment, they generate scenarios in which an anthropomorphic robot initiates and ensures joint attention. Results show that a robot's joint attention behavior is judged as having better performance in an interactive task and is perceived as natural behavior.

During human-robot collaboration, some actions generated by robots are perceived as intentional. This motivates work aiming to produce more natural and more human-like robot behaviors and movement trajectories. Indeed, [37] conduct an experimental study on the smoothness and precision of a humanoid robot's gestures, finding that people react more quickly to smooth gestures compared to abrupt "machine-like" gestures. In [38] the authors add another factor of human-like movement, which is variability. To avoid repetitive movements, the algorithm adds variance by exploiting redundant spaces and underutilized spaces of the input movement, allowing the creation of multiple movements from a single input. Their algorithm ensures that the robot can satisfy the task constraint while maintaining the variance and human quality of the movement.

To conclude on the concept of analysis, understanding, and communication of intention in HRI. It concerns representing, understanding, and reasoning about human collaborators' intentions through a capacity called theory of mind (ToM), with the objective of making the robot's intention as transparent as possible for the human allowing a human partner to easily deduce the targeted goal or purpose of its action.

On the other hand, there are many approaches to intention understanding that have no basis in ToM and rather focus on representing activity with probabilistic models that allow early prediction of action completion. This sub-domain is still in its early stages and many opportunities remain in understanding human intentional action. Therefore, enhanced use of context and multiple modalities emerge as important avenues still to be explored in depth.

E. Social Navigation: To properly plan a path through an environment, the robot must perform dynamic perception of the environment and localize its position as well as that of its goal in the environment map. The robot will detect and avoid obstacles reactively and incorporate newly detected obstacles through trajectory planning algorithms. This strategy works well for obstacles, such as an object on the ground that was not present at the last pass, but humans are another type of obstacle. People in an environment are there for a certain purpose, and their behavior is determined by their intention. They may want to interact with the robot and the robot may have the objective of interacting with them. Thus a social robot must treat navigation as a social planning problem, and to do so, it must reason about its own intentions with respect to human intentions in the environment. This coordination of intentions is what makes the problem intrinsically different compared to reactive obstacle avoidance. Indeed, a robot wanting to interact must first detect the desired person for interaction, and then approach them in a safe and appropriate manner while communicating its intention to interact.

In [39] the authors present an in-depth study of behavioral theories and concepts on a robot's social navigation. They classify research on social navigation into several categories: proxemics and human approach [40], [41] unfocused interaction, such as passing in the corridor and avoidance; targeted interaction,

such as dialogue, person following [42] and side-by-side walking [43].

In [44] the authors propose a movement model for a robot arm according to simple movement rules with the aim of reducing stress when a robot is near a human, and must move more slowly.

The authors in [44] suggest the notion of Human-Aware Motion Planner (HAMP), which takes into consideration the human context. They define a "safety grid" and a "visibility cost grid" that depend on human orientation and posture (sitting/standing). The cost-minimizing planner then finds the navigation path most suited to the human. In this same context, [46] develop a HAMP planner by adding harmonious rules such as no collisions, no interference, waiting, etc. taking into account not only human position and orientation, but also human movement vector. Furthermore, the motion planner considers social hierarchies and passage priorities in narrow spaces.

The particular behavior of approaching humans has been widely studied by the research community, who have focused on how a robot should pass by a person when in a narrow space, like a corridor. [47] reference social distances, but especially find that people prefer the robot to make as wide a detour as possible around them.

The authors in [48] model the social navigation problem as a set of constraints to be weighted and optimized simultaneously. These constraints include: minimizing distance traveled during obstacle avoidance, avoiding the person (including certain social aspects, such as passing on the right and respecting personal space), and maintaining constant speed.

In [49] the authors propose a method to generate a smooth path taking into account a number of sensor readings and dynamically evaluated rules. Approaching humans in public spaces poses additional problems related to the environment and the set of people in the scene. In this context, [50] use a heuristic approach in combination with a Support Vector Machine (SVM) classifier to model the territory of a shop in a shopping center with the aim of allowing robots to find appropriate positions to approach customers.

There is also research combining social navigation with verbal instruction and dialogue [51]. This allows robots to have more detailed context about the space in which they navigate, navigate according to explicit human intentions, and engage humans in dialogue about the navigation they are attempting.

As this domain develops, combining social navigation with non-verbal behaviors like haptics in an HRI scenario may present a promising research direction. Similarly for social navigation of aerial vehicles, given that there is very little research dealing with this theme, we can therefore expect more work in this domain in the coming years.

F. Human Perception: For a social robot, perceiving humans for interaction is a fundamental challenge for many HRI systems. The requirements of HRI applications distinguish this challenge from that of the general domain of computer vision. Indeed, perception in the context of HRI requires dynamic, incremental, real-time, semantically rich, and often multiparty interaction [52].

In HRI literature, there are three general aspects of human perception: recognizing human poses [53] and [54] and their characteristics, analyzing their actions [55], [56] and detecting engagement for interaction [57].

|53] propose a description of humans with features coding the angular relationship between their body parts. A feature vector is then mapped to a code-word in Hidden Markov Models (HMM) and used for real-time recognition of whole-body gestures in an HRI context.

While previous works are all focused on human search and tracking, others try to determine specific human attributes that could impact interaction. [58] detect human age and gender based on their movement. Their system uses image and video processing, 2D Discrete Wavelet Transform and 2D Fast Fourier Transform to extract human movement characteristics. Then the AdaBoost algorithm is used to classify gender and age based on spatio-temporal information.

In recent years, with the launch of the RGB-D depth camera and its ready-to-use skeletal tracker, more and more research has used it.

Kahlouche et al. [59] used data acquired by a Kinect sensor for activity recognition as inputs to a deep learning architecture. Their approach has been used to interact with a service robot and achieved satisfactory accuracy on a set of six interaction activities (Hello, Call; Stop; Pointing, Approaching, and move away).

Sensor fusion is another approach to make activity recognition more robust in a real environment. For example, [60] used verbal descriptions as additional input alongside video for activity recognition. This helps disambiguate situations that are visually similar.

Hand gestures are a particularly important aspect of social interaction. [61] used sensor fusion (depth and RGB) to track hands in real-time in a cluttered environment. A Monte-Carlo based update process reduces false positives

and an online skin color learning algorithm adapts to different skin colors, clothing, and lighting conditions. To encourage HRI-focused activity recognition, [62] published a Dataset composed of 18 activities captured by the RGBD camera for human-robot interaction. The dataset includes both communicative and non-communicative actions.

In summary, the rich literature on human perception for human-robot interaction is quite mature and builds on the long tradition of computer vision. Human perception methods in recent years focus on three challenges. The first generally consists of detecting people and their characteristics; this includes locating them; estimating their poses, genders, and ages; detecting their faces and extracting gaze information. The second classifies activities in real-time once a person is detected and characterized. This is often done using skeleton tracking and sensor fusion methods. Finally, to successfully use activity recognition capacity, a social robot must also know which person is attempting to interact with it and if they are still interacting. This motivates a third challenge of perception in HRI: detecting initial and continuous engagement.

3. CONCLUSION

Social robots as those for which people apply a social model to understand and interact with them. These robots are not designed to replace humans but to be social partners that contribute to human cognitive well-being. In this paper, a state of the art on Human-Robot Interaction has been presented. We have seen some examples of social robots that can be guide tour, receptionist, waiters, or mediators in society. The HRI paradigm is presented which is actually constituted of several factors notably: the level of robot autonomy; the nature of exchanged information, the organization of the human-robot team, and the type of communication (proximate or remote).

We have also presented taxonomy on research work in the interdisciplinary domain of HRI where each underlying axis can be considered as the application and extension of robotics research in a socially interactive context. For example, robot navigation is seen through a new vision of social navigation, both for human social needs and expressing social behaviors during navigation. Similarly, for robot perception techniques that have been adapted and extended to specifically perceive information useful for social interaction. This is valid for all fundamental capabilities related to robotics, such as expression of verbal and non-verbal behaviors, design of emotional models; reasoning about intentions, collaboration, and learning.

References

[1] KAHLOUCHE, S. « Implémentation de techniques avancées d'Interaction Homme Robot pour un robot de service » (Doctoral dissertation, Ecole Nationale Supérieure d'informatique), 2023.

[2] Scholtz, J. (2003). Theory and evaluation of human robot interactions. 36th Annual Hawaii International Conference on System Sciences, 2003. Proceedings of The, 10 pp.

[3] Fernandez, R., Cheng, P., Smith, B., Fenton, T., Boraey, Y., Nhlabatsi, A., ... & Fetais, N. (2023, October). Interactive Graphical Access Control Tools. In 2023 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC) (pp. 255-259). IEEE.

[4] Wee, C., Yap, K. M., & Lim, W. N. (2021). Haptic interfaces for virtual reality: Challenges and research directions. IEEE access, 9, 112145-112162.

[5] Thomaz, A., Hoffman, G., & Cakmak, M. (2016). Computational Human-Robot Interaction. Foundations and Trends in Robotics, 4(2–3), 104–223.

[6] Zhang, S., Wei, Z., Nie, J., Huang, L., Wang, S., & Li, Z. (2017). A Review on Human Activity Recognition Using Vision-Based Method. Journal of Healthcare Engineering, 2017.

[7] Tellex, S., Gopalan, N., Kress-Gazit, H., & Matuszek, C. (2020). Robots That Use Language. Annual Review of Control, Robotics, and Autonomous Systems, 3(1), 25–55.

[8] Read, R., & Belpaeme, T. (2016). People Interpret Robotic Non-linguistic Utterances Categorically. International Journal of Social Robotics, 8(1), 31–50.

[9] Torrey, C., Fussell, S. R., & Kiesler, S. (2013). How a robot should give advice. 2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI), 275–282.

[10] Kirk, N. H., Nyga, D., & Beetz, M. (2014). Controlled Natural Languages for language generation in artificial cognition. 2014 IEEE International Conference on Robotics and Automation (ICRA), 6667–6672.

[11] Salem, M., Ziadee, M., & Sakr, M. (2013). Effects of Politeness and Interaction Context on Perception and Experience of HRI (pp. 531–541).

[12] Bremner, P., & Leonards, U. (2015). Speech and Gesture Emphasis Effects for Robotic and Human Communicators. Proceedings of the Tenth Annual ACM/IEEE International Conference

[13] Admoni, H., Weng, T., Hayes, B., & Scassellati, B. (2016). Robot nonverbal behavior improves task performance in difficult collaborations. 2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI), 51–58

[14] Van den Bergh, M., Carton, D., De Nijs, R., Mitsou, N., Landsiedel, C., Kuehnlenz, K., Wollherr, D., Van Gool, L., & Buss, M. (2011). Real-time 3D hand gesture interaction with a robot for understanding directions from humans.

2011 RO-MAN, 357–362.
[15] Liu, P., Glas, D. F., Kanda, T., Ishiguro, H., & Hagita, N. (2013). It's not polite to point Generating socially-appropriate deictic behaviors towards people. 2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI), 267–274.
[16] Ou, S., & Grupen, R. (2010). From manipulation to communicative gesture. 2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI), 325–332.
[17] Droeschel, D., Stückler, J., & Behnke, S. (2011). Learning to interpret pointing gestures with a time-of-flight camera. Proceedings of the 6th International Conference on Human-Robot Interaction - HRI '11, 481.
[18] Yamazaki, A., Yamazaki, K., Kuno, Y., Burdelski, M., Kawashima, M., & Kuzuoka, H. (2008). Precision timing in human-robot interaction. Proceeding of the Twenty-Sixth Annual CHI Conference on Human Factors in Computing Systems - CHI '08, 131.
[19] Chen, T. L., & Kemp, C. C. (2010). Lead me by the hand: Evaluation of a direct physical interface for nursing assistant robots. 2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI), 367–374.
[20] Knight, H., Toscano, R., Stiehl, W. D., Chang, A., Wang, Y., & Breazeal, C. (2009). Real-time social touch gesture recognition for sensate robots. 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems, 3715–3720.
[21] Flagg, A., & MacLean, K. (2013). Affective touch gesture recognition for a furry zoomorphic machine. Proceedings of the 7th International Conference on Tangible, Embedded and Embodied Interaction - TEI '13, 25.
[22] Silvera-Tawil, D., Rye, D., & Velonaki, M. (2014). Interpretation of Social Touch on an Artificial Arm Covered with an EIT-based Sensitive Skin. International Journal of Social Robotics, 6(4), 489–505.
[23] Ammi, M., Demulier, V., Caillou, S., Gaffary, Y., Tsalamlal, Y., Martin, J. C., & Tapus, A. (2015). Haptic Human-Robot Affective Interaction in a Handshaking Social Protocol. ACM/IEEE International Conference on Human-Robot Interaction, 2015-March, 263–270.
[24] Bevan, C., & Stanton Fraser, D. (2015). Shaking Hands and Cooperation in Tele-present Human-Robot Negotiation. Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction, 247–254.
[25] Ham, J., Cuijpers, R. H., & Cabibihan, J.-J. (2015). Combining Robotic Persuasive Strategies: The Persuasive Power of a Storytelling Robot that Uses Gazing and Gestures. International Journal of Social Robotics, 7(4), 479–487. https://doi.org/10.1007/s12369-015-0280-4
[26] Xue, T., Wang, W., Ma, J., Liu, W., Pan, Z., & Han, M. (2020). Progress and prospects of multimodal fusion methods in physical human–robot interaction: A review. IEEE Sensors Journal, 20(18), 10355-10370.
[27] Norman, D. A. (2004). Emotional design. Ubiquity, 2004(January), 1–1.
[28] Breazeal, C., Buchsbaum, D., Gray, J., Gatenby, D., & Blumberg, B. (2005). Learning From and About Others: Towards Using Imitation to Bootstrap the Social Understanding of Others by Robots. Artificial Life, 11(1–2), 31–62.
[29] Hirth, J., Schmitz, N., & Berns, K. (2011). Towards Social Robots: Designing an Emotion-Based Architecture. International Journal of Social Robotics, 3(3), 273–290
[30] Lang, C., Wachsmuth, S., Hanheide, M., & Wersing, H. (2013). Facial communicative signal interpretation in human-robot interaction by discriminative video subsequence selection. 2013 IEEE International Conference on Robotics and Automation, 170–177
[31] Alazrai, R., & Lee, C. S. G. (2012). Real-time emotion identification for socially intelligent robots. 2012 IEEE International Conference on Robotics and Automation, 4106–4111. https://doi.org/10.1109/ICRA.2012.6224587 on Human-Robot Interaction, 255–262.
[32] Sanghvi, J., Castellano, G., Leite, I., Pereira, A., McOwan, P. W., & Paiva, A. (2011). Automatic analysis of affective postures and body motion to detect engagement with a game companion. Proceedings of the 6th International Conference on Human-Robot Interaction - HRI '11, 305.
[33] Gopnik, A., Sobel, D. M., Schulz, L. E., & Glymour, C. (2001). Causal learning mechanisms in very young children: Two-, three-, and four-year-olds infer causal relations from patterns of variation and covariation. Developmental Psychology, 37(5), 620–629. https://doi.org/10.1037/0012-1649.37.5.620
[34] Huang, C.-M., & Mutlu, B. (2016). Anticipatory robot control for efficient human-robot collaboration. 2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI), 83–90.
[35] Trafton, J. G., Hiatt, L., Harrison, A., Tanborello, F., Khemlani, S., & Schultz, A. (2013). ACT-R/E: An Embodied Cognitive Architecture for Human-Robot Interaction. Journal of Human-Robot Interaction, 2(1), 30–55.
[36] Pandey, A. K., & Alami, R. (2010b). Mightability maps: A perceptual level decisional framework for co-operative and competitive human-robot interaction. 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems, 5842–5848.
[37] Riek, L. D., Rabinowitch, T.-C., Bremner, P., Pipe, A. G., Fraser, M., & Robinson, P. (2010). Cooperative gestures: Effective signaling for humanoid robots. 2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI), 61–68.
[38] Gielniak, M. J., Liu, C. K., & Thomaz, A. L. (2013). Generating human-like motion for robots. The International Journal of Robotics Research, 32(11), 1275–1301.
[39] Rios-Martinez, J., Spalanzani, A., & Laugier, C. (2015). From Proxemics Theory to Socially-Aware Navigation: A Survey. International Journal of Social Robotics, 7(2), 137–153.
[40] Admoni, H., Dragan, A., Srinivasa, S. S., & Scassellati, B. (2014). Deliberate delays during

robot-to-human handovers improve compliance with gaze communication. Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction, 49–56.

[41] Patompak, P., Jeong, S., Nilkhamhang, I., & Chong, N. Y. (2020). Learning Proxemics for Personalized Human–Robot Social Interaction. International Journal of Social Robotics, 12(1), 267–280. .

[42] Brookshire, J. (2010). Person Following Using Histograms of Oriented Gradients. International Journal of Social Robotics, 2(2), 137–146.

[43] Murakami, R., Morales Saiki, L. Y., Satake, S., Kanda, T., & Ishiguro, H. (2014). Destination unknown. Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction, 471–478.

[44] Hanajima, N., Goto, T., Ohta, Y., Hikita, H., & Yamashita, M. (2005). A motion rule for human-friendly robots based on electrodermal activity investigations and its application to mobile robot. 2005 IEEE/RSJ International Conference on Intelligent Robots and Systems, 3791–3797.

[45] Sisbot, E. A., Marin-Urias, L. F., Alami, R., & Simeon, T. (2007). A Human Aware Mobile Robot Motion Planner. IEEE Transactions on Robotics, 23(5), 874–883.

[46] Chi-Pang, L., Chou, C.-T., Chiang, K.-H., & Fu, L.-C. (2011). Human-Centered Robot Navigation—Towards a Harmoniously Human–Robot Coexisting Environment. IEEE Transactions on Robotics, 27(1), 99–112.

[47] Pacchierotti, E., Christensen, H., & Jensfelt, P. (2006). Evaluation of Passing Distance for Social Robots. ROMAN 2006 - The 15th IEEE International Symposium on Robot and Human Interactive Communication, 315–320. https://doi.org/10.1109/ROMAN.2006.314436

[48] Kirby, R., Simmons, R., & Forlizzi, J. (2009). COMPANION: A Constraint-Optimizing Method for Person-Acceptable Navigation. RO-MAN 2009 - The 18th IEEE International Symposium on Robot and Human Interactive Communication, 607–612.

[49] Pandey, A. K., & Alami, R. (2010a). A framework towards a socially aware Mobile Robot motion in Human-Centered dynamic environment. 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems, 5855–5860.

[50] Satake, S., Iba, H., Kanda, T., Imai, M., & Saiki, Y. M. (2014). May i talk about other shops here? Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction, 487–494.

[51] Van den Bergh, M., Carton, D., De Nijs, R., Mitsou, N., Landsiedel, C., Kuehnlenz, K., Wollherr, D., Van Gool, L., & Buss, M. (2011). Real-time 3D hand gesture interaction with a robot for understanding directions from humans. 2011 RO-MAN, 357–362. https://doi.org/10.1109/ROMAN.2011.6005195

[52] Xue, T., Wang, W., Ma, J., Liu, W., Pan, Z., & Han, M. (2020). Progress and prospects of multimodal fusion methods in physical human–robot interaction: A review. IEEE Sensors Journal, 20(18), 10355-10370.

[53] Yang, H.-D., Park, A.-Y., & Lee, S.-W. (2007). Gesture Spotting and Recognition for Human–Robot Interaction. IEEE Transactions on Robotics, 23(2), 256–270.

[54] Shotton, J., Fitzgibbon, A., Cook, M., Sharp, T., Finocchio, M., Moore, R., Kipman, A., & Blake, A. (2011). Real-time human pose recognition in parts from single depth images. CVPR 2011, 1297–1304.

[55] Vieira, M., Faria, D. R., & Nunes, U. (2016). Real-time application for monitoring human daily activity and risk situations in robot-assisted living. Advances in Intelligent Systems and Computing, 418, 449–461.

[56] Coronado, E., Villalobos, J., Bruno, B., & Mastrogiovanni, F. (2017). Gesture-based robot control: Design challenges and evaluation with humans. 2017 IEEE International Conference on Robotics and Automation (ICRA), 2761–2767.

[57] Lee, J., Kiser, J. F., Bobick, A. F., & Thomaz, A. L. (2011). Vision-based contingency detection. Proceedings of the 6th International Conference on Human-Robot Interaction - HRI '11, 297.

[58] Handri, S., Nomura, S., & Nakamura, K. (2011). Determination of Age and Gender Based on Features of Human Motion Using AdaBoost Algorithms. International Journal of Social Robotics, 3(3), 233–241

[59] Kahlouche, S., Belhocine, M., & Menouar, A. (2021). Real-time human action recognition using deep learning architecture. International Journal of Computational Intelligence and Applications, 20(04), 2150026. (TOG), 43(4), 1-12.

[60] Teo, C. L., Yang, Y., Daume, H., Fermuller, C., & Aloimonos, Y. (2012). Towards a Watson that sees: Language-guided action recognition for robots. 2012 IEEE International Conference on Robotics and Automation, 374–381.

[61] McKeague, S., Liu, J., & Yang, G.-Z. (2013). An Asynchronous RGB-D Sensor Fusion Framework Using Monte-Carlo Methods for Hand Tracking on a Mobile Robot in Crowded Environments (pp. 491–500).

[62] Chrungoo, A., Manimaran, S. S., & Ravindran, B. (2014). Activity Recognition for Natural Human Robot Interaction (pp. 84–94).

.
.
.