Foundations and Trends® in Robotics
Vol. 4, No. 2-3 (2013) 105–223

DOI: 10.1561/2300000049

Computational Human-Robot Interaction

Andrea Thomaz
The University of Texas at Austin USA
athomaz@ece.utexas.edu

Guy Hoffman
Cornell University, Ithaca USA
hoffman@cornell.edu

Maya Cakmak
University of Washington, Seattle USA
mcakmak@uw.edu

Contents

Abstract

We present a systematic survey of computational research in human-robot interaction (HRI) over the past decade. Computational HRI is the subset of the field that is specifically concerned with the algorithms, techniques, models, and frameworks necessary to build robotic systems that engage in social interactions with humans. Within the field of robotics, HRI poses distinct computational challenges in each of the traditional core research areas: perception, manipulation, planning, task execution, navigation, and learning. These challenges are addressed by the research literature surveyed here. We surveyed twelve publication venues and include work that tackles computational HRI challenges, categorized into eight topics: (a) perceiving humans and their activities; (b) generating and understanding verbal expression; (c) generating and understanding non-verbal behaviors; (d) modeling, expressing, and understanding emotional states; (e) recognizing and conveying intentional action; (f) collaborating with humans; (g) navigating with and around humans; and (h) learning from humans in a social manner. For each topic, we suggest promising future research areas.

A. Thomaz, G. Hoffman and M. Cakmak. *Computational Human-Robot Interaction*. Foundations and Trends$^{®}$ in Robotics, vol. 4, no. 2-3, pp. 105–223, 2013.

DOI: 10.1561/2300000049.

1

Introduction

The field of human-robot interaction (HRI) is expanding and maturing. At the time of writing, dedicated publications on HRI and social robotics research include two special-interest journals and three conferences, in contrast to a single conference and no dedicated journals in 2005. In addition, HRI is a research topic which is increasingly solicited and included in the broader robotics community.

The goal of this survey paper is to provide a systematic overview of the field of HRI over the past decade (from 2005 to 2015), with a focus on the computational frameworks and algorithms currently used to enable robots to interact with humans. Two influential surveys of the field were published in 2003 and 2007 [Fong et al., 2003, Goodrich and Schultz, 2007], and a book chapter surveyed part of the HRI literature in 2008 [Breazeal et al., 2008]. This survey starts roughly where Goodrich and Schultz [2007] left off, covering what has proven to be the most active period of HRI research thus far.

This paper's focus, however, is different from the previous surveys. As the research area has developed, we have identified a lack of a systematic survey focusing specifically on computational HRI research. This subfield of HRI, which includes algorithmic and systems-oriented work is distinct from the large body of research dealing with the empirical, psychological, cultural, and user-interface aspects of the field. So far, there has not been a comprehensive survey article covering computational HRI. In addition, to the best of our knowledge, there has never been a systematic review of the literature in an attempt to represent the bibliometric trends, balance, and distribution of work in HRI. This paper aims to fill these gaps.

1.1 Methodology

While no survey paper can argue for exhaustiveness, we employed a systematic methodology when selecting for inclusion. Our search covered the entire archive of the top-rated journals and refereed conference proceedings which publish work on HRI and social robotics. This included traditional robotics journals and conferences, one human-computer interaction conference, and specialized HRI and social robotics venues. In total, we surveyed twelve venues:

- IEEE Transactions on Robotics (T-RO)
- International Journal of Robotics Research (IJRR)
- Autonomous Robots (AuRo)
- Journal of Human-Robot Interaction (JHRI)
- International Journal of Social Robotics (IJSR)
- Robotics: Science and Systems (RSS)
- International Conference on Robotics and Automation (ICRA)
- International Conference on Intelligent Robots and Systems (IROS)
- International Conference on Human-Robot Interaction (HRI)
- International Symposium on Robot and Human Interactive Communication (RO-MAN)

- International Conference on Social Robotics (ICSR)
- ACM Conference on Human Factors in Computing Systems (CHI)

For these twelve venues, we considered the entire archive published since January 2005 and selected papers based on pre-defined inclusion criteria, described in the following section.

1.1.1 Inclusion Criteria

Delineating the research which contributes to the technologies underlying socially interactive robots is a non-trivial question of field boundary and demarcation. With an eye on the grand challenge of building autonomous socially intelligent robots, our goal was to specifically cover computational, i.e., algorithmic and robotics-oriented (as opposed to psychology-oriented), and synthetic (as opposed to descriptive or inferential) research. This excludes all user studies only measuring human responses to robot behavior or designs. Of the computational papers considered, we further limited the survey by including only work that has a clear element of robotics and a clear element of social interaction.

In other words, our rule-of-thumb for inclusion requires that both the social and the computational should be present in the research, and that the intended application of the work is in robotics. To formalize this, we defined several inclusion and exclusion criteria, organized by type and topic of the research papers we considered:

- **Perception of Humans** — There is a large body of work in the robotics and HRI literature concerned with the perception of humans. Out of those we include only the subset of papers in which the perception was geared towards, or focused on, social interaction. We either exclude or only briefly mention work that is aimed at detecting and tracking people in the environment generally, without specific application to HRI, such as perception for situational awareness or context understanding.

  There are a number of venues concerned with computational perception, such as the Conference on Computer Vision and Pattern Recognition (CVPR) and the International Conference on

Computer Vision (ICCV), to name two. The fact that we did not survey these venues inherently narrows our scope to research aimed at robotics applications and at HRI in particular. This means that we do not survey some of the core computational perception work, even though it has undoubtedly affected the field of HRI significantly.

- **Learning** — Machine learning also constitutes a large part of robotics research. We focus on the subset of papers in which learning happens either with an eye on social interaction or directly through social interaction. We do not include work merely treating human data as a learning database for inference, even if it is geared toward robotics.

  A similar point can be made for foundational work in machine learning as we made earlier with respect to computational perception. Research in venues such as the International Conference on Machine Learning (ICML) or Neural Information Processing Systems (NIPS) is not represented in this survey, even though much of it has clear relation to the work discussed herein.

- **Collaboration, Navigation, and Manipulation** — In human-robot collaboration, navigation, and manipulation papers, we focus on those that include a distinctly social aspect. This means that we exclude a large body of efficiency-centric collaborative robotics work found in industrial robotics research. We do include a few selected works on collaborative manipulation, in particular those that relate to intentionality.

- **Autonomy** — As a rule, we include only research in which the robot has at least some autonomy, or that is concerned with developing methods that serve robot autonomy. This excludes most, if not all, work with the Wizard-of-Oz (WoZ) methodology, with a few exceptions, described below.

We cannot claim that the boundaries of this survey are crisply delineated. In fact, it would be fair to say that more papers were borderline for inclusion than clear-cut. For example, we include some purely empirical studies which are designed with computational questions in mind, or have clear implications for autonomously interactive robotic systems. We include such work in particular when it helps frame the discussion of subsequent computational research.

Overall, we identified, read, and considered 926 papers out of the original several thousands of papers published in the above-mentioned venues in the survey time frame. Our criteria narrow this list even further, resulting in a total of 375 papers representing the state of the art in computational HRI.

1.2 Overview

HRI is an interdisciplinary field with roots and connections in several more established disciplines of robotics and computer science. This is reflected in the categorization of the work surveyed here. Each section can be viewed as the application and extension of robotics research to the socially interactive context.

For example, techniques from the field of robot perception have been adapted and extended to specifically perceive information used for social interaction, and in particular to reason about human intention. Similarly, whereas the broader field of robotics studies kinematics and motion planning, a socially interactive robot needs to view these issues in the context of nonverbal communicative behavior. Motion planning is made socially aware in order to communicate intents and create bonds. The broader topic of machine learning for robotics gives rise to research in socially-guided robot learning, building on human models of tutelage and instruction. Similarly, the long tradition of robot navigation is seen through a new lens of social navigation, both accounting for human social needs and expressing social signals during navigation.

Inspired by this perspective, Figure 1.1 shows an overview of this paper. The paper flows from fundamental robot capabilities, such as perception of human activities, expression of verbal and nonverbal behavior, and the role of emotion models in HRI, to higher-level social robot skills, including reasoning about intentions, collaboration, navigation, and learning.

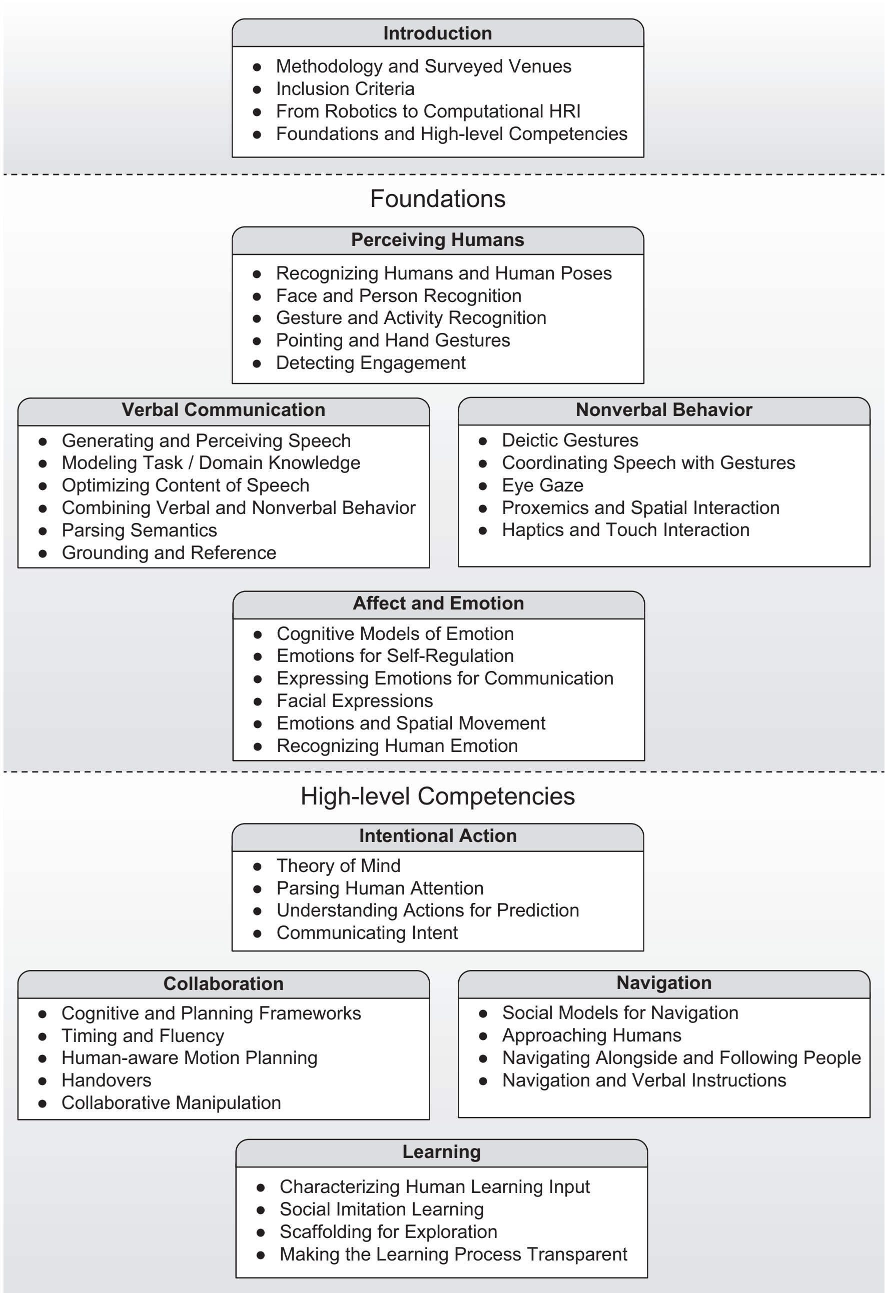
Introduction
Methodology and Surveyed Venues
Inclusion Criteria
From Robotics to Computational HRI
Foundations and High-level Competencies
Foundations
Perceiving Humans
Recognizing Humans and Human Poses
Face and Person Recognition
Gesture and Activity Recognition
Pointing and Hand Gestures
Detecting Engagement
Verbal Communication
Generating and Perceiving Speech
Modeling Task / Domain Knowledge
Optimizing Content of Speech
Combining Verbal and Nonverbal Behavior
Parsing Semantics
Grounding and Reference
Nonverbal Behavior
Deictic Gestures
Coordinating Speech with Gestures
Eye Gaze
Proxemics and Spatial Interaction
Haptics and Touch Interaction
Affect and Emotion
Cognitive Models of Emotion
Emotions for Self-Regulation
Expressing Emotions for Communication
Facial Expressions
Emotions and Spatial Movement
Recognizing Human Emotion
High-level Competencies
Intentional Action
Theory of Mind
Parsing Human Attention
Understanding Actions for Prediction
Communicating Intent
Collaboration
Cognitive and Planning Frameworks
Timing and Fluency
Human-aware Motion Planning
Handovers
Collaborative Manipulation
Navigation
Social Models for Navigation
Approaching Humans
Navigating Alongside and Following People
Navigation and Verbal Instructions
Learning
Characterizing Human Learning Input
Social Imitation Learning
Scaffolding for Exploration
Making the Learning Process Transparent

Figure 1.1: Overview of the Paper Structure

1.2.1 Foundations

Sections 2–5 cover basic capabilities and modalities of interaction. These core behaviors are precursors to successful interaction with humans.

The first of these skills, covered in Section 2, is the ability to perceive humans in a social context. The computational issues arising from this goal include a number of challenges: First, a robot might need to recognize a human social partner, find their face, and possibly recognize their identity. Then, a robot could recognize gestures, track the focus of the human's attention, identify activities, and detect the human's engagement or disengagement with the robot.

Next, Section 3 covers systems and methods aimed at generating and understanding verbal expression, geared toward human dialog with social robots. This includes a variety of technical challenges, including optimizing speech content, expressing task and domain knowledge, understanding context, and grounding verbal content in the physical world. We also cover work that looks at paralinguistics, such as the tone of voice (vocalics) and the timing of speech acts.

Section 4 considers nonverbal behavior. To support social interaction, robots need the ability to generate and understand the variety of nonverbal behavior exhibited in human communication. This includes the detection and generation of body movements (kinesics), pointing gestures, speech-accompanying gestures, and gaze, as well as space and territory management (proxemics), and touch interactions (haptics).

A central aspect of human nonverbal behavior is the perception and generation of emotional behaviors and signals. Work on this topic, covered in Section 5, skirts the boundary between Affective Comput-ing and HRI, including computational models representing emotional states and the use of emotion for robot self-regulation. We also cover frameworks and methods for generating emotion expression and techniques for detecting human emotional states in the context of human-robot social interaction.

1.2.2 High-level Competencies

In the second portion of this survey we discuss social behaviors that build on the skills covered in the first portion. This begins with the expression and recognition of intentional behavior. Humans have natural tendency to parse the world on intentional boundaries. Therefore, understanding, predicting and reasoning about intentions is fundamental to interaction. Section 6 surveys work concerned with the automatic detection, classification, and recognition of human intention. This includes work on Theory of Mind capabilities for robots, on the prediction of human activities as intentional agents, and on mechanisms to achieve joint attention. This section concludes by looking at how robots can generate actions that communicate intent in an appropriate way, based on animation principles and legibility optimization. The capacity to understand and communicate intentional action then serves as the basis for the last three sections covering the social behaviors of collaboration, navigation, and learning.

Section 7 includes research focused on human-robot collaborative activities, a highly active subfield of computational HRI. In order to collaborate successfully with a human, a robot needs to adjust its motion planning algorithms, optimizing for social aspects of the movement. A large body of work deals with computational challenges in embodied shared activities, including collaborative planning and scheduling, while others consider timing, anticipation, and team fluency. Finally we look at two highly-studied instances of human-robot collaboration: object handovers and collaborative manipulation of a shared object.

All of the above sections are equally applicable to stationary and mobile robots. However, mobile robots have unique challenges associated with the social aspects of their use of space. Section 8 surveys research on socially-aware robot navigation and mobility. In many ways this is a particular case of collaborative behavior. First, mobile robots need to recognize and generate intentional behavior. Then, there are social aspects of the navigation itself, including approaching people, moving around people, and accompanying humans along their walking path.

Finally, Section 9 looks at machine learning in the context of HRI, including robot learning guided by humans. This capacity also builds on the robots' ability to represent and reproduce intentional behavior in order to help human teachers give better instruction. The section covers the particular features of human-generated machine learning input as well as human-inspired learning techniques, such as scaffolding. In this section, we stress the importance of social signals in robot learning, enabling a more transparent learning process by the robot.

2

Perceiving Humans for Social Interaction

In order to interact with a person socially it is necessary to perceive them, both to initiate the interaction and then to maintain it. As a result, the perception of humans for social interaction is a fundamental computational challenge for many HRI systems. A social robot has to find people in the environment, distinguish between different people, parse their actions and activity, and estimate their engagement.

The application requirements of HRI distinguish this computational challenge from that of the general field of Computer Vision, or more broadly Computational Perception. Perception in an HRI context requires parsing dynamic, incremental, real-time, semantically rich, and often multi-party interactions. Databases of images and video, which are the cornerstone of Computational Perception research, have little in common with scenarios of a human interacting with a robot.

To support the goal of perceiving humans, HRI systems use a combination of sensors and modalities, including vision, audio, and touch. The reader is encouraged to refer to Yan et al. [2014] for a structured review of sixteen social robots, organized by their sensing capabilities, the applications they serve, and the perception tasks they achieve.

In the HRI literature, we found three overarching topics of computational human perception: recognizing humans and their features, parsing their actions, and detecting engagement for interaction.

2.1 Recognizing Humans: Features, Faces, and Gaze

The first precursor to a robot's ability to interact with a human is the ability to recognize humans and understand aspects of them that may influence the interaction. This has been a topic of HRI research for many years, with much of the recent work in the field dealing with estimating human poses and activities, detecting their face and facial features, and determining their gaze.

Several pieces of work are devoted to recognizing humans and their body pose from a static vantage point. McColl et al. [2011] use a combination of a thermal camera and a time-of-flight (ToF) range camera to segment the head and lower arms and detect their configuration, even in the presence of intersecting body parts. Yang et al. [2007] propose describing a human with features encoding the angular relationship between their body parts. A feature vector is then mapped to a codeword of Hidden Markov Models (HMMs) and used for real-time recognition of whole body gestures in an HRI context. Ho et al. [2005] also use HMMs, but to model and recognize a human's motion. Temporal Difference (TD) Learning is applied to adjust the tracking performance online, allowing the system to track and recognize hand motions of a walking person.

For mobile robots to interact, they have to track a person's position and angle with respect to them. This has often been achieved with the use of laser range finders [Svenstrup et al., 2009]. In one example, Panangadan et al. [2010] track the movement of people in both indoor and outdoor environments using laser range finders. The tracking data are then segmented into sequences arising from distinct activities by representing the tracks as probability distributions. This allows for the classification of these activities. Luber and Arras [2013] detect and learn socio-spatial relations between individuals and are able to track group formations using 2D range data.

While the above works are all focused on finding and tracking humans, others try to determine specific attributes about the human that might impact an interaction. Handri et al. [2011] detect the age and gender of humans based on their motion. Their system uses image and video processing, 2D Discrete Wavelet Transformation and 2D Fast Fourier Transformation to extract human motion features. Then, the AdaBoost algorithm is used for classifying gender and age based on spatio-temporal information. Carcagnì et al. [2014] present a different approach for gender prediction in real time. They make use of commercial face tracking software and use an SVM to classify gender.

In addition to tracking and recognizing the poses and attributes of a human in general, a social robot can use face recognition for the purpose of recognizing a particular human, in order to have a user-specific interaction. Again, in contrast to the core face recognition research community, the papers surveyed here are concerned with real-time recognition in realistic conditions of interactive robotic deployments. For example, Aryananda [2009] describes an incremental and unsupervised face recognition system that was evaluated in an eight-day-long experiment in which a robot autonomously detected, tracked, and segmented face images during spontaneous interactions with over 500 passersby.

Computational methods for HRI face recognition vary. Hanheide et al. [2008] links interactive introduction of interlocutors with an online learning face classification scheme based on active appearance models. In Pateraki et al. [2009], a mobile robot finds humans by locating and tracking faces and facial features using Least Squares Matching (LSM), an approach that can overcome the problems of variable scene illumination and the head in and off-plane rotations. Raducanu and Dornaika [2010] track and recognize facial expressions exploiting facial action parameters estimated by an appearance-based 3D face tracker. The complexity of the non-linear facial expression space is modeled through a manifold, the structure of which is learned using Laplacian Eigenmaps. The projected facial expressions are then recognized based on a Nearest Neighbor classifier.

Related to face tracking, a social robot can benefit from interpreting the gaze direction of a human interlocutor. Michalowski and Simmons

[2006] track people and their gaze using vision and laser range data. Their approach classifies people using a categorical model of attention: present (far from the robot), attending (idling closer to the robot), engaged (next to the robot), and interacting (actively participating in an exchange with the robot). Work related to attention recognition will be discussed more in detail in Section 6.2 in the context of establishing joint attention in a human-robot interaction.

2.2 Activity and Gesture Recognition

Interactive scenarios for social robots require not only the ability to find and track humans, but also to understand their gestures and actions. One aspect of this challenge is the recognition of whole-body activities and gestures. Prior to the availability of high-accuracy low-cost skeleton tracking using commercial depth sensors, a large portion of the research was devoted to motion and pose tracking from traditional sensors. Jenkins et al. [2007], for example, perform monocular tracking and action recognition for movement imitation from partial observations at interactive rates.

In recent years, out-of-the-box skeletal trackers using RGB-D camera systems are used most often. Anjum et al. [2014] uses a Kinect sensor as input for an activity recognition classifier, with a basic SVM classifier that achieves near perfect accuracy on a closed set of eight activities. Sensor fusion is another approach for making recognition of activities in real environments more robust. For example, Teo et al. [2012] uses language descriptions as additional input alongside video for activity recognition. This helps disambiguate situations that are visually similar. Burger et al. [2012] use HMMs to recognize gestures for robot commands, fusing this with speech recognition. In Droeschel et al. [2011b], a combination of two laser range finders (LRFs) with one RGB camera and one time-of-flight camera are employed to detect joint attention nonverbal gestures. The LRFs suggest candidates for humans and the camera verifies these candidates. Both cameras are used to detect gestures, including gaze and pointing gestures. For the latter, the authors use HMMs modeling the various stages of a deictic point.

Hand gestures are a particularly important aspect of social interaction and are addressed separately in computational HRI. Using a monocular camera, Chuang et al. [2011] use a Bag-of-Words method to detect and recognize hand posture based on a so-called Appearance and Relative Position Descriptor in combination with a spectral embedding clustering algorithm. They then use Continuously Adaptive Mean Shift (CAMshift) to track hand motion in real time. McKeague et al. [2013] use sensor fusion (depth and RGB) to track hands in crowded environments in real time. A Monte-Carlo update process reduces false positives and an online skin color learning algorithm copes with varying skin color, clothing, and illumination conditions.

To encourage HRI-focused activity recognition, Chrungoo et al. [2014] released an annotated RGBD human-robot interaction dataset consisting of 18 unique activities including ten stylized gestures and eight conventional activities of daily living. The dataset includes both communicative and non-communicative actions.

2.3 Detecting Engagement

The above-mentioned activity and gesture recognition research presupposes that the robot knows the human is trying to interact with it. However, consider a robot in an office lobby or a shopping mall. Such a robot should not assume that every human it detects wants to interact. Hence, another challenge for social robots is to detect the human's interaction engagement, both initially and in an ongoing fashion. Ideally this should be framed as a more generic recognition problem than that of recognizing a particular gesture or activity. This way the person does not have to make a specific gesture or use a speech command to initiate an interaction.

In Feil-Seifer and Matarić [2005], a robot selects a person for interaction based on the perceived desire of humans to interact with it. They use a multimodal approach, tracking legs with a laser range finder, and gestures with a camera, then supplementing these sensors with speech recognition. Finke et al. [2005] use sonar sensors to recognize human movements making the assumption that people interested in interacting will approach the robot. The robot distinguishes objects from humans by assuming that only people move by themselves and a Hidden Markov Models approach detects the approach movements correctly in approximately 80% of the experimental cases.

As opposed to recognizing a particular behavior, Lee et al. [2011] use human-robot contingency as an indication of engagement. Using motion-based features, their method can classify when a nearby human has made a contingent response to the robot's actions with 79% accuracy. Lee et al. [2012] additionally incorporate an audio signal, and Chu et al. [2014] demonstrate this multimodal contingency detection model working in real-time.

In a reverse approach, Torta et al. [2012] study how to attract the attention of a human, comparing different modalities (eye contact, blinking, waving, speaking). They find that sound generates the fastest reaction time, whereas trying to establish eye contact is slowest. In Torta et al. [2015] the authors present a follow-up study that shows that only using speech is the fastest way to attract a human's attention, even compared to a multimodal approach of using speech and actions combined.

Once engagement is initiated, Rich et al. [2010] propose a computational model and score for recognizing strength of continued engagement by combining four types of events: directed gaze, mutual gaze, conversational adjacency pairs, and back channel communication.

In summary, research in computational methods for perceiving humans in recent years focuses on three challenges. The first is generally detecting people and their features. This includes finding them; estimating their pose, gender and age; detecting their faces and extracting gaze information. Once a person is detected and characterized, a robot can classify their activity in real time for interaction. This is often done using skeleton tracking and sensor fusion methods. Finally, in order to successfully use the capability of activity recognition, a social robot has to also know which person is attempting to interact with it and whether they are still interacting. This motivates a third area of computational HRI perception research: detecting initial and ongoing engagement.

The rich literature of perceiving humans for human-robot interaction is quite mature and builds on a longer tradition of Computer Vision and Computational Perception. Most of the work on parsing the perceived human's activity focuses on single persons and a natural extension for future research is to parse activities of more than one person interacting with the robot. This includes the perception of people who are not interacting with the robot, but are of interest to the robot's performance. Furthermore, while much field-tested research has been conducted in public settings, there has been little non-laboratory work on the perception of humans in more private environments, such as the home or the office.

3

Verbal Communication in Social Robots

Verbal communication is a key aspect of human social behavior and is therefore critical for human-robot interaction. Speaking enables the transfer of semantically rich information between the robot and its human interlocutor. The general challenge of speech recognition and production has resulted in vast areas of research in the general computer science, artificial intelligence (AI), and human-computer interaction (HCI) literature. In this section, we focus on the computational methods and systems specific to verbal communication for social robots.

Language is often related to application and context. Over the years, speech systems have been used in HRI across a large number of scenarios, including cooking instruction [Torrey et al., 2007], task directions [Namera et al., 2008], storytelling [Al Moubayed et al., 2009], city exploration [Weiss et al., 2010] and tour guidance [Shiomi et al., 2010]. There are verbally communicative robots for snack delivery [Lee et al., 2010], in hospitals [Raman et al., 2013], for furniture assembly [Tellex et al., 2014], in the home [Lohse et al., 2008], and as receptionists [Salem et al., 2013b].

What all of these systems have in common is that the content of the robot's language is tied to its physical embodiment and environment.

Speech acts need to coordinate with the robot's physical movement in space (whether that movement is functional or expressive). At the same time the language used by the robot should correctly ground spatial references to the robot's environment. These considerations are reflected in the research surveyed below.

Regarding the asymmetry of language parsing and production: In humans, language recognition and generation processes are tightly coupled. In robots, generating speech—once the robot has a string of text—is relatively easy with text-to-speech engines. Recognizing speech is much more challenging, as it involves additional uncertainty on the input channel (noise, accents, timing) beyond the semantic content. As a result, many robots speak but do not understand spoken language. For example, the robot in Torrey et al. [2007] uses text-to-speech to guide the user through a task, but the human types in their input; similarly, Weiss et al. [2010] use speech and image as output modalities, but gesture and touch screen as inputs. This asymmetry can be an issue in the social relationship between humans and robots, as people might expect a social robot that generates speech to be able to recognize speech of similar complexity. In this section, we cover these two aspects of speech separately: first the generation of verbal behavior, then its recognition.

3.1 Generating Verbal Behavior

There are two primary computational challenges in generating verbal behavior for social robots: (1) generating the content of the speech and (2) generating the delivery parameters of the speech. The first challenge includes modeling the task and the interaction domain in order to produce the content of speech. The second challenge consists of paralinguistic properties, coordinating with nonverbal behavior, and the overall timing of the speech acts with respect to the robot's other actions.

3.1.1 Modeling Domain Knowledge for Speech Production

Speech content depends on the task the social robot is performing and the domain or context in which it is intended to perform. Researchers have employed a variety of frameworks to model this knowledge for speech production. Examples include employing a customized variant of the Artificial Intelligence Mark-up Language (AIML) for question-answering [Torrey et al., 2007], or using Attempto Controlled English (ACE), a subset of standard English with restricted lexicon, syntax and semantics, formally described by a set of construction rules. Kirk et al. [2014] use the latter to paraphrase incomplete or ambiguous natural language instructions and ask questions about missing information to complete the task.

Often language can be tailored to the task at hand. For example, the verbal production system in the Okuno et al. [2009] shopping mall robot generates verbal route directions using a "skeletal description" of sentences, each involving an action (e.g., go straight) relative to a landmark (e.g., a distinct building), for example "turn left at the bank".

3.1.2 Determining the Content of Speech

A key difference between human-robot dialog versus general AI dialog systems is that producing the correct speech content in an HRI scenario is dependent on the physical surroundings, the collaborative situation as detected by a variety of sensors, and the human's spatial behavior.

In the context of human-robot collaborative tasks, Tellex et al. [2013] created a system to generate clarification questions for disambiguation, based on an information theoretic strategy. In Tellex et al. [2014], the robots ask for help when a failure is detected during a furniture assembly task. The help request is optimized for ease of understanding by a human listener using "inverse semantics", which uses a probabilistic model of the human's process for interpreting a request. Knepper et al. [2015] also make use of the inverse semantics approach to provide natural language error explanation with the aim of asking humans for help in an otherwise autonomous robot task planner. This produces utterances that describe the action required from the human to complete the task. Their system models the probability that the human will understand the request.

Explanation and clarification of robot behavior have ties to work in which natural language commands given by humans are parsed into robot controllers and verified for realizability. In this context, Raman et al. [2013] describe a system that uses linear temporal logic to generate provably correct controllers and introduce a method for generating natural language phrases explaining the causes of failure to the human when their instruction is unrealizable.

In another example, St Clair and Mataric [2015] use explanatory language for the purpose of coordinating a collaboration. Combining human activity recognition and a communication planner, the system reasons about the role the human is taking in the collaboration and then produces either a self-explanatory feedback or a role reallocation feedback in order to avoid conflicting actions. Elaborations can also be dynamically generated. Torrey et al. [2007] present a system that provides verbal elaborations only if the user appears to need it, as indicated by gazing at the robot or through slow progress on the task.

Generating accurate and efficient referring expressions is an important capability for interactive robots. Fully-elaborated grounded spoken references are needed when there are no assumptions of shared knowledge between the two parties or the ability to use other modalities. However, in human-robot interactions, referring expressions are often supported with nonverbal behaviors (e.g. pointing gestures).

Fang et al. [2015] propose an incremental collaborative model for generating referring expressions. Rather than generating a single utterance that uniquely specifies a target item in the environment, the robot iteratively constructs *installments* that reduce the number of target candidates and, along the way, confirms that the user understands what the robot is referring to. Their method for generating referring expressions involves empirically learning weights over features of objects to be included in the expression and whether or not to use a pointing gesture as part of the referring expression. Foster et al. [2008] generate multi-modal referring expressions that use *haptic-ostensive references*, that is, references which involve manipulating an object being referred to, as opposed to just gestural-deictic references (e.g., pointing). This includes narrating actions as they are performed by the robot, correcting a human partner's actions by handing them a different object, or referencing objects that are already in hand by shaking them.

The content of speech can also be selected based on the persona the robot is intended to convey. Salem et al. [2013b, 2014] manipulated the politeness of speech in a direction-giving receptionist robot based on socio-linguistic theories of “face”. A robot that is *bald on record* would not try to minimize threats to the listener’s face; e.g., in response to being asked directions to an unknown destination it would say “I have absolutely no idea”. In contrast, a robot that employs *positive politeness* would attempt to imply shared desires; e.g., it would respond “Sorry, I don’t know where it is because I am new”.

3.1.3 Paralinguistics

Paralinguistics are properties of spoken verbal behavior that are embedded into the verbal message. Based on research on politeness and informal speech in humans, Torrey et al. [2013] propose ways to make advice-giving robots sound less commanding. This involves making the robot use *hedges* (“I think", “maybe”, “kind of") or *discourse markers* (“you know,” “I mean,” “well,” “just,” “like,”) as part of their directions.

Aly and Tapus [2013] developed a system that transforms an input utterance to a new utterance augmented with gestures reflecting specified personality traits (e.g., introverted-extroverted). Based on the desired robot personality traits, the verbal response can vary in verbosity, polarity, and repetitions, while the associated gesture can vary in amplitude, direction, rate, and speed. Niculescu et al. [2013] varied robot voice pitch to manipulate the personality of a robot (extrovert or introvert), and Chao and Thomaz [2013] showed that manipulating turn-taking parameters, with in particular timing, changes the social dynamics of an interaction, making a robot seem more passive or active.

Paralinguistics can also be used to make robots more persuasive. Andrist et al. [2013] propose to do so by including different cues of expertise in the robot’s verbal recommendations. That includes five cues associated with rhetorical ability: *good will*, *prior experience*, *organization*, *metaphors*, and *fluency*. Nakagawa et al. [2013] study the effect of speech volume for a robot that gives advice during a mentally demanding task. They find that using a small voice is not different from using normal voice; however, *whispering* appears to increase motivation and results in improved task performance. Chidambaram et al. [2012] generate verbal and nonverbal cues of persuasiveness based on the literature on human persuasiveness, including tone, proximity, gaze, and loudness.

In some cases, starting with Breazeal's Kismet (2004), robots produce *only* paralinguistics, in the form of tonally-controlled giberrish or sound effects ("non-linguistic utterances" or NLU). Read and Belpaeme [2014] study the use of NLUs (expressive beeps) for communication. They demonstrate that the context in which the sound is made most strongly determines people's interpretation of what the sound means. However, certain *iconic* sounds, like a rising single tone, have well established meanings that might not be overwritten by context. Chao and Thomaz [2013] demonstrate that robot-generated gibberish speech can evoke a natural turn taking dialog, and Fischer et al. [2014] suggest that a robot that passes a human with a beep is preferred over a silent robot and that a rising pitch contour is preferred over a falling one.

Another aspect of successful paralinguistics is the naturalness of the verbal expression. Text to speech systems were originally developed for screen readers and were based on monologue speech. As a result they can sound unnatural as part of a human-robot dialog. To mitigate this problem Sugiura et al. [2014] developed a speech synthesizer that uses recordings of non-monologue speech.

Other work combines paralinguistics with other nonverbal channels: Bremner and Leonards [2015] add emphasis to parts of a spoken verbal message, for disambiguation or salience, by using pitch accents in the intonation and by adding beat gestures, such as a downward vertical hand movement timed to coincide with the emphasized word. Al Moubayed et al. [2009] models filled pauses, gestures (head nod, head shake), facial expressions (smile, eyebrow movement), and acoustic prominence to accompany utterances in a storytelling scenario. This work is rooted in a long tradition of virtually embodied conversational agents and the coordination of speech and illustrator gestures [Cassell, 2000].

3.1.4 Coordinating Verbal and Nonverbal Behavior

Much of the integration of verbal production with nonverbal behavior deals with the timing and coordination aspects between the two modalities for emphasis, clarity, and naturalness. To support the design of these coordinated systems, Shi et al. [2010] developed a markup language that allows for easy programming of communicative behaviors that incorporate nonverbal behaviors with utterances. The programmer of the robot writes sentences tagged with meta-data about desired nonverbal behaviors such as gazes, gestures, and standing points. Their system automatically fills in details of the nonverbal behaviors.

Namera et al. [2008] use human-human communication data to develop a system that learns the timing of nonverbal behavior, such as nodding, and action response, such as grasping an object, in relation to a confirmation utterance generated in response to a task command. The robot in Okuno et al. [2009] gives verbal directions augmented with deictic gestures to help people find their ways to different shops inside a mall. Salem et al. [2013a] develop a closed-loop approach for synchronizing arm gestures with speech. An empirically learned forward model roughly estimates the timing of gestures and a feedback loop adapts the onset of speech based on progress of the gesture.

Nodding is a particularly common nonverbal gesture that happens during speech. Several works address the timing of head nods [Sidner et al., 2006] or head turns [Yamazaki et al., 2008] with respect to speech. Ishi et al. [2010], for example, present a model to generate nods during a dialog based on rules inferred from observations of human-human dialogs. These include nodding to express agreement, but also in synchrony with speech at the last syllable of a phrase or in strong phrase boundaries.

To manage the conversational floor with multiple humans, Mutlu et al. [2009a] suggested a system whereby the robot uses gaze cues to establish the participant roles (addressee, bystander, overhearer). They designed gaze cues for each role based on theories of human social communication and formal observations of human-human interactions. Using their model, the robot was able to establish roles by varying whether and how long the robot gazed at people during greetings, turn exchanges and the core of the conversation.

3.2 Recognizing Verbal Behavior

Understanding verbal expression in a social human-robot dialog also faces unique challenges, but can be aided by taking advantage of the situated nature of the interaction, which can help frame and contextualize the recognition problem.

A precursor to the topic of understanding verbal expressions is the observation that people talk differently to robots than they do to computers. This is in part due to the cultural aspects of conversing with a robot, but also because humans have multi-modal access to the robot's agency, beyond mere text production. Lexical entrainment (also referred to as alignment) is people's tendency to adopt the terms of their interlocutor. Iio et al. [2009] demonstrated lexical entrainment in people during interaction where they instruct the robot to move objects. People both adopted specific terms and types of terms from the robot. Lohse et al. [2008] found that people adapt their input utterances and gestures when the robot expresses that it did not understand a previous utterance.

In the rest of this section, we review the literature concerned with the understanding of verbal expression along two categories: understanding the semantic content of the utterances and understanding its relationship to the physical world (grounding and referring).

3.2.1 Parsing Semantics

The semantic parsing problem for social robots is more specific than a general language understanding challenge and is often concerned with converting speech into robot control, using the task at hand as a context.

Some pieces of work are geared towards programming a robot using verbal instruction. Miller et al. [2007] use a context-free grammar modeled as a Dynamic Bayes Net, which is compiled into a Hidden Markov Model to provide verbal programming of a robot's action sequences.

Ralph and Moussa [2008] parse commands for moving a manipulator, suggesting ways to bridge the infinite continuous action space with the discrete command space.

In the system developed by Deits et al. [2013], people can use arbitrary natural language to command a robot. They use a Generalized Grounding Graph [Tellex et al., 2011], which assigns grounding probabilities to speech components and uses an entropy measure to find the most ambiguous part of the command. The robot then attempts to clarify that part by using yes-no, “What do you mean by . . . ”, and “Can you rephrase . . . ” questions. The human’s answer is then merged into the grounding graph to produce the appropriate action.

Raman et al. [2013] devised a system to command robots to perform complex high-level tasks using natural language by converting natural language specifications into Linear Temporal Logic formulas that are used for synthesizing controllers. Matuszek et al. [2010] present a method based on data-driven machine translation to convert natural language spatial directions to a robot navigation path on a known map. Similarly Kollar et al. [2010] present a method that computes the most likely path from spatial description clauses extracted from natural human directions. A spatial description clause, such as “go past the computers,” involves a verb, a landmark, and a spatial relation to the landmark, which can be associated with actions taken as part of a path on a map.

Other probabilistic and likelihood-based approaches include Howard et al. [2014], who infer the most likely set of planning constraints (i.e., a subregion of the state space) from natural language instructions, such as “move near the red box and the blue crate"; and Fasola and Mataric [2014], who use a probabilistic chaining approach to interpret spatial language instruction sequences combining recency and spatial grounding criteria for anaphora resolution. MacGlashan et al. [2015] ground natural language commands into reward functions rather than actions, using demonstrations of a command being carried out in the environment.

Expanding the robot’s vocabulary during the interaction, Cantrell et al. [2011] can utilize natural language explanations to teach new tasks identified by an unknown verb. For example, the explanation "To follow someone means to stay within one meter of them" gives the robot the ability to correctly respond to the previously unknown command "follow me" based on its existing understanding of staying within a certain distance of a target.

3.2.2 Understanding Referring and Grounded Speech

Just as a social robot can use its own gestures and gaze to refer to objects in space and to ground elements of speech, it can use a variety of cues to help understand human spoken language. These include the environment, the context, and the human's nonverbal behaviors.

Guadarrama et al. [2013] use visual and spatial information in conjunction with the semantic parsing of utterances to interpret either direct reference to objects, or references through objects' spatial relationships, and to execute commands. This capability is compositional: the robot can understand complex commands that refer to multiple objects, their relationships, and actions. In Kollar et al. [2010], a robot's path is generated from natural human direction, by grounding landmarks and spatial relationships. Landmark phrases are grounded in the perceptual frame of the robot based on large databases of tagged images from the Web. Also using images, Blisard and Skubic [2005] model spatial referencing terms such as *front, behind, left, right, between* in 2D images in order to enable commanding the robot using these references. In Hemachandra et al. [2014], the robot learns semantic maps from a combination of human descriptions and its own perceptual information.

Howard et al. [2014] overcome the exponential scalability problem of grounding language directly into robot actions by grounding language instead into a number of planning constraints, and using a trajectory planner under those constraints to find the correct actions for the language instruction.

In some cases visual perspective-taking is used to disambiguate references [e.g., Berlin et al., 2006]. Ros et al. [2010] and Lemaignan et al. [2012] extend that work by representing the robot's knowledge as an ontology and are able to identify ambiguous referents on a shared table situated between the human and the robot. Humans can make statements informing the robot of new facts, give orders to the robot, and ask questions on declarative knowledge. Other systems combine verbal commands with deictic gestures [e.g., Brooks and Breazeal, 2006]. In one example, Fransen et al. [2007] use gestures to understand prepositional referents. The robot acts if the gesture unambiguously determines the referent. If multiple alternatives are possible the system will ask "Which one?" or "Where?", and will process human clarification.

Clarification is a powerful tool that a socially interactive robot can use to help speech understanding. In this context, Sattar and Dudek [2011] combine cost, including safety risks and action costs, with a confidence measure of having correctly understood the human instruction based on the recognition HMM. These two factors decide whether to comply with the instruction or trigger a clarification request from the human. In Sattar and Little [2014], the authors extend this model by allowing the robot to clarify, verify, and possibly discard individual subcommands in a certain task, based on their risk and potential cost. Cantrell et al. [2010] use incremental semantic parsing of verbal instructions in order to provide early back-channel feedback to human speakers. They also suggest ways to discard disfluencies ("like", "um", etc.) and use corrections ("I mean ...", "no, actually") as overrides in order to improve the understanding process.

A final challenge is the separation of speakers in multi-human-robot interaction. Valin et al. [2007] suggest a system that can tease apart simultaneous speech in a mobile robot. They use the Geometric Source Separation algorithm with a microphone array and assumed known source locations. Gomez et al. [2012] use a speech-recognition criterion to learn the parameters of traditional signal processing methods to enable multi-party human-robot interaction with distant talkers. Utilizing a multi-modal approach, Trafton et al. [2008] combine auditory sound source localization with vision based human tracking in order to determine the speaker during a multi-speaker conversation.

In summary, verbal communication is a fundamental communication channel for social robots to use with human partners. Generating verbal behavior has computational challenges around determining both the content and the delivery parameters of what to say. Research in the HRI literature on generating speech has focused on how to represent knowledge for the purpose of speech production, on algorithms for optimizing speech content (particularly for explanations and referring expressions), and on algorithms to generate the paralinguistic cues and timing parameters which determine how speech is delivered. Research in speech understanding for HRI also tackles not only the semantic aspects of interacting with a robot, but also the possibilities and advantages of grounded and referred speech. A social robot understands language in the physical context it is embedded in, and can use the human's nonverbal behavior to help understand their verbal utterances.

In many ways robots today are able to generate more sophisticated speech than they are able to recognize, and this asymmetry is an important anchor for future opportunities in this space. In contrast to the number of works on generating paralinguistic cues, there is very little research into how these cues can be recognized and used in understanding verbal behavior. In addition, incremental algorithms for the recognition of verbal behavior need to be studied and developed further as this skill is essential for the dynamic nature of human robot dialogs. Moreover, a promising area for future research is the interplay of generation and understanding. The recognition and generation of verbal utterances is an ongoing tightly coupled process of reaching common ground through dialog. Computational HRI can make use of this coupling to advance both speech production and speech understanding in social robots.

4

Communicating with Nonverbal Behavior

Nonverbal behavior is a well-studied area of human behavior, with roots leading back to the 19th century, most notably to Darwin's "The Expression of Emotions in Man and Animals" [Darwin, 1873]. Contemporary textbooks on the topic [e.g., Moore et al., 2013, Knapp et al., 2013] agree for the most part on the categories of nonverbal behaviors: body movements and gestures (kinesics), including facial expressions and eye gaze; managing space and territory (proxemics); touch (haptics); tone of voice (vocalics); and appearance, including morphology, clothing, and body alterations.

To support social interaction, robots have to control and understand these modalities of nonverbal behavior. As a result, nonverbal behavior has been an active area of research since the beginning of the field of HRI. In this section, we give an overview of both the generation and recognition aspects of nonverbal behavior, broken down by categories and modalities of nonverbal behaviors.

HRI research has given unequal attention to the various modalities of nonverbal behavior, with the bulk of the research concentrated around kinesics, and an additional emphasis within kinesics on gaze, gestures, and facial expressions. As a result, the following section, too, is heavily focused on kinesics research. There is slightly less research on computational proxemics, mostly in the context of social navigation and much less research on computational haptics.

Nonverbal behavior is closely tied to other HRI capacities. For example, proxemics has a clear overlap with social navigation and will be discussed in more detail in Section 8. Similarly, haptics partially overlaps with the work on collaborative manipulation covered in Section 7.5. Most of the vocalics literature was covered in Section 3.1.3 above. Facial expressions, due to their strong connection to the communication of emotions, are discussed in more detail in Section 5 on emotional communication in HRI. Finally, research addressing appearance is usually considered part of the robot design literature and not included in this survey.[1]

4.1 Categories of Kinesics

A bulk of the computational HRI kinesics research is influenced by seminal work by Ekman and Friesen on gestures and facial expressions [Ekman and Friesen, 1969], as well as the work of Argyle on gaze and eye-contact [Argyle and Dean, 1965, Argyle et al., 1973].

Ekman and Friesen identify five major categories of kinesics: *emblems*, which are symbolic gestures replacing speech elements, such as the ‘thumbs up’ gesture; *illustrators*, which are gestures of all kinds accompanying speech acts; *affect displays*, expressing emotion, primarily through facial expressions; *regulators*, which are nods, hand movements, postures, and gaze behaviors that help coordinate conversation; and *adaptors*, such as rubbing one’s face, shifting hair, or tapping on the table. Illustrators are a large group of nonverbal behavior, so Ekman and Friesen further subdivide them into a number of categories, including *deictic*, or pointing, gestures, *batons*, which are rhythmic movements

[1] It is worth mentioning the extentive work done on nonverbal behavior expression in the virtual characters and embodied conversational agent literature, which address some of the same issues as discussed in this section. Note that in this survey we only discuss nonverbal behavior expression in robotics, which has distinct parameters and challenges, related to the embodied, situated, and physically constrained nature of the nonverbal behavior performed by robots.

emphasizing speech, and other gestures to depict ideas, delineate space, and imitate human activities.

This categorization is useful for—and used by—HRI researchers, and we, too, use this taxonomy in the following section. In kinesics, we cover deictic illustrators, regulators and batons, and then move to discuss gaze and eye contact, proxemics, and haptics.

4.2 Deictic Gestures

Deictic gestures are a kind of illustrator used to "point to a present object", i.e., an object in the space of the pointer. Robots need to be able to accurately interpret human pointing and generate legible and disambiguated pointing on their own.

Brooks and Breazeal [2006] present a framework to recognize deictic gestures from a human. A spatial reasoning system taking into account the hand-eye relationships and the environment resolves pointing and object references. This work additionally uses speech cues to constrain the recognition problem. In recent years, the recognition problem has been significantly simplified by readily available human pose trackers using RGB-D cameras. Van den Bergh et al. [2011] present a real-time pointing detection system using a Microsoft Kinect sensor for a robot giving directions. Similarly, Perez Quintero et al. [2013] achieve pointing recognition with a Kinect for object selection.

Numerous studies investigate aspects of robots using pointing gestures to communicate with humans. These are shown to increase people's information recall [Huang and Mutlu, 2013] as well as task performance and perceived workload [Lohse et al., 2014]. Eye gaze is shown to significantly assist the recognition of pointing gestures [Iio et al., 2010, Häring et al., 2012], and in fact, Liu et al. [2013] present a study that shows that people do not usually point to refer to a person, but instead use eye gaze or casual pointing. A robot is seen as more polite when it balances the two types of pointing in a social situation.

Deictic gestures to the same object can be designed in a variety of ways. Sauppé and Mutlu [2014] compare several such variations on a small humanoid robot, including pointing, presenting, touching, and sweeping, and find their effectiveness to be strongly related to the context, including the object's distance from the referrer and proximity to other objects. In non-humanoid robots, Williams et al. [2013] show that the direction and pitch of the robot's head was important. People seem to interpret this as eye gaze, and expect it to be coordinated with the pointing gesture.

With the aim of automatically generating deictics and other kinesics, Ou and Grupen [2010] use a learning approach for a robot to acquire communicative behaviors from pragmatic actions. Using hierarchical Markov Decision Processes and a set of primitive behaviors (including tracking visual features and reaching for and manipulating objects), the robot learns that a failed reach communicates pointing to the human and that appropriately adding eye gaze improves performance. In another learning-based system, Droeschel et al. [2011a] extract a set of body features from depth and amplitude images of a time-of-flight (ToF) camera and train a model of pointing directions using Gaussian Process Regression.

4.3 Regulators and Batons: Coordinating Gesture with Speech

In some cases gestures accompany speech for rhythmic emphasis (*batons*) and in others to coordinate verbal communication (*regulators*). Salem et al. [2012] present a baton generation system for the Honda ASIMO robot based on a motor control program that works to schedule and align speech and gesture appropriately. People interacting with the robot rate interactions with gesture as better than speech-only, but congruent vs. incongruent scheduling has little impact.

Yamazaki et al. [2008] present a system for a museum robot to move its head at "interactionally significant" points of an explanation, such as at transition points, together with deictic words, in response to a question, upon keywords, or with unfamiliar words. The approach is shown to positively affect human engagement. In early work on robot regulators, Sidner et al. [2006] build a robot that recognizes head nods from a human conversation partner, and can generate response nods. Experimental results were unexpected: People nod at the robot if it is talking, regardless of whether or not the robot recognizes their nods or responds appropriately. However, Hashimoto et al. [2007] show that a robot's "speaker's nod" using a motion model adapted from human timing, angle, and velocity data can lead to greater human-like rating and emotional expression of interlocutors compared to a randomly nodding robot.

4.4 Eye Gaze

In their seminal work, Argyle et al. identify different communicative functions of human gaze, with some similarity to Ekman's kinesic categories [Argyle and Dean, 1965, Argyle et al., 1973]. They argue that gaze is used to signal interpersonal attitudes, to illustrate speech for emphasis and context, to regulate dialog and turn-taking, and to create or avoid intimacy. Research in computational HRI is influenced by this categorization and is working to develop systems that generate similar social gaze behaviors. Since gaze detection, tracking, and interpretation is a vast literature in the more general HCI community and not specific to HRI, we do not survey it in this section.

Studies show humans to be influenced in a variety of ways by robot gaze. Staudte and Crocker [2009] demonstrate that a human's own gaze behavior and understanding of the robot's speech content is modulated by the coordination of that robot's speech and gaze. They also show that people's eyes follow the robot's gaze. Admoni et al. [2013] find that people are more accurate at recognizing shorter, more frequent fixations than longer, less frequent ones. In a collaborative task, people are also found to take spatial and context clues from brief (400ms) robotic glances [Mutlu et al., 2009b]. Other work in the storytelling domain, however, shows people's ability to retain story information to be only influenced by gestures, and not by gaze [Van Dijk et al., 2013].

Gaze cues can be particularly useful during robot-to-human object handovers. Admoni et al. [2014] find that these cues influence people's compliance with the direction indicated by the gaze in ambiguous handover situations. The addition of deliberate-seeming delays increases compliance also in non-ambiguous situation, even when this results in counterintuitive human behavior. Moon et al. [2014] show that people reach for an offered object earlier when a robot signals via gaze to the handover target location. A field study shows gaze to be a significant cue to identify the receiver of a handover in a large group setting [Kirchner et al., 2011].

Eye gaze can be used by robots to help manage the conversational floor. For example, looking away can signal cognitive effort, helping to regulate a conversation. Based on this insight, Andrist et al. [2014] present an autonomous three-function gaze control system to control a robot. The robot uses face-tracking to engage in mutual gaze, idle head motion to increase lifelikeness, and purposeful gaze aversions to achieve regulatory conversational functions. In contrast, to achieve more naturalistic and engaging gaze behavior, Sorostinean et al. [2014] present a social attention system that tracks a person but attends to strong motion when detected in its visual field. In a non-human-inspired take on the expression of robotic gaze attention, Yamaguchi and Hashimoto [2009] add an LED color cue to a pan-tilt robot head, mirroring the color of the object the camera is looking at.

Gaze can also be used by robots to signal interpersonal attitudes and to create intimacy and trust. A humanoid robot's gaze has a positive impact on trust for difficult human decisions, and it also improves participants' task performance on easy trials but hinders it on difficult trials [Stanton and Stevens, 2014]. Furthermore, robots are found to be more persuasive when they use gaze [Ham et al., 2015]. However, we did not find computational systems that make use of eye gaze to modulate interpersonal attitudes or intimacy.

In order to generate a realistic robotic eye gaze behavior, Kuno et al. [2006] analyze human head-orientation data in a museum setting. Based on this, they present a system for a guide robot. The robot alternates gaze between exhibition items and human audiences (through face detection) while explaining the exhibits. In a variation on human-like gaze, Matsumaru et al. [2005] employ gaze-like behavior in a robot using a glass ball with an eyeball projected on it. Their mobile robot communicates direction of motion using horizontal placement, and speed of motion by changing the eyeball "openness".

4.5 Proxemics

Proxemics refers to the socially communicative or expressive aspects of spatial positioning and orientation. Social robots can infer human intention and make predictions based on human proxemics and can use proxemics rules to generate more socially appropriate behaviors. Note that much of generative proxemics is covered in Section 8 on social navigation.

Mead et al. [2011] suggest spatial metrics for analyzing human behavior, in combination with other measures such as voice loudness. They demonstrate the feasibility of autonomous real-time annotation of these metrics during multi-person social encounters using a sensor suite including an overhead camera, a markerless motion capture system, and an omnidirectional microphone. In Mead et al. [2013] the authors extend this to show the impact of individual vs. physical vs. psychophysical features in recognition of proxemic social behaviors. The psychophysical features build more successful Hidden Markov Models for this recognition task (72% vs. 56% accuracy). Using a Gaussian Mixture Model with the naïve-Bayes approach, also from overhead camera observations, Feil-Seifer and Mataric [2011] achieve a 91.4% accuracy rate in classifying task specific behaviors in robot-child-parent interaction, and demonstrate that these classes are sufficient for distinguishing between positive and negative reactions of the child toward the robot.

Tasaki et al. [2005] focus on the spatial relationships between a robot and multiple people during interaction. Using tactile sensing, face detection, and sound localization, the robot estimates the distance between humans, and maintains a “friendliness space map”, leading to robot attention behavior selection.

On the generative end of the spectrum, early work by Brooks and Arkin [2007] combines proxemics, emblematic gestures, postures, head pose, orienting, leaning, and head nods in a behavioral overlay framework. They also use a social hierarchy model to generate appropriate behaviors for different classes of human interlocutors, such as an old friend, a stranger, or an authority figure. In another work aimed at generating behavior from proxemic information, Yamaoka et al. [2009] observe people’s proxemic behavior in joint attention situations and develop a model of behavior for robots to detect a partner's attention shift and appropriately adjust their own body position and orientation in establishing joint attention with the partner.

4.6 Haptics

Work in robot touch as a nonverbal communicative channel is sparser than kinesics, gaze, and proxemics, both for touch detection and touch generation. Several robotic systems infer human intentions through force sensing on the robot's end effectors—e.g., Chen and Kemp [2010], where the robot "feels" the human pull it and infers the intended motion vector and speed, or Ferland et al. [2013] which uses joint-space impedance control of the arms' differential elastic actuators. This approach has numerous variants in collaborative manipulation, as expanded upon in Section 7.5.

Researchers have also explored touch sensors for robotic skins, with algorithms classifying types of affect intentions. A seminal example is the "Huggable" robot skin [Stiehl et al., 2005, Knight et al., 2009], and the artificial fur developed by Flagg and MacLean [2013]. In recent work, Silvera-Tawil et al. [2014] design a stretchable sensitive skin and a classifier based on the LogitBoost algorithm to classify six emotions and six social messages transmitted by humans when touching an artificial arm.

People have been shown to be influenced by a robot actively touching them, for example to persist in a monotonous task [Nakagawa et al., 2011]. Most studies, however, focus on the specific iconic touch of a *handshake*. For example, Ammi et al. [2015] show that stiffness of joints and grasp force can modulate the sense of dominance a robot emits, and Bevan and Stanton Fraser [2015] finds that handshaking before negotiations improves cooperation. Zeng et al. [2012] present a generative model for handshaking. Their hybrid reactive/deliberative model tries to maintain its own position trajectory, but is also influenced by the human, resulting in trajectories similar to human-human handshakes.

In summary, nonverbal behavior is one of the most studied fields in HRI, with much of the research focusing on a subset of behaviors, notably on deictic gestures, some illustrators, and eye gaze. Surveying the literature of systems that produce and recognize nonverbal behavior, the bulk of computational HRI research deals with recognizing pointing gestures and with generating batons and regulators. We did not find work on emblems, adaptors, and other illustrators, such as the ones that depict actions, ideas, or spatial relationships. Surprisingly, while there is a considerable amount of research on the importance and effects of eye gaze, there is effectively little computational research on the production of appropriate gaze behaviors in HRI.

Overall, in a survey of ten years of computational HRI research, we find a consistent trend: Compared to the vast amount of empirical work on nonverbal behavior dominating the psychological subfield of HRI, the computational literature on the topic is sparser than expected. For example, our survey indicates that there is little work on the recognition of illustrators, or on the production of deictic gestures. Most other kinesics types are effectively ignored. Even eye gaze, which is a cornerstone of empirical HRI, does not have a similarly rich computational literature. Most of the known functions of gaze are not being utilized in the design of computational systems. The topic of HRI haptics, or interacting with robots through touch, is also ripe for future research, especially in the context of robots touching humans, or in the exploration of new materials for haptic HRI. These gaps all offer promising areas of future research in nonverbal behavior for computational HRI.

5

Affect and Emotion in Social Robots

Of the range of communicative and regulatory roles of nonverbal behavior, the expression and recognition of emotions stand out and play a major role. Emotions are a cornerstone of how humans communicate effectively with each other and come to infer another person's mental state as it relates to the current context or situation. Indicative of this, the relationship between emotions and nonverbal behavior receives separate and extended attention in seminal texts on human nonverbal behavior [e.g., Darwin, 1873, Ekman and Friesen, 1969]. The connection between technology and emotions is later emphasized by Norman [2004], who argues that emotional interaction should play a central part in the design of technological artifacts as well, particularly in robot design.

Affective interaction with robots falls within the broader research field of affective computing [Picard, 1997]. Both Norman and Picard argue for deep underlying models of emotions in affective technology, rather than surface level interactions in which the technology merely uses emotions as a communication technique. In the field of computational HRI, we find work spanning this spectrum between underlying affect models on the one end and surface-only emotional interactions on the other.

A seminal example of an emotionally expressive social robot is Kismet [Breazeal and Scassellati, 1999, Breazeal, 2004]. Kismet's attention system includes a motivation model including *drives* and *affects*. Each drive has a desired operating point and the dynamic state of the drives influences a subsequent 3-dimensional affective state. Socially communicative behaviors are triggered by affect states with the terminal goal of keeping the internal drives in their desired state. This dynamic plays out in the robot's nonverbal behavior, resulting in a continuous nonverbal social exchange with a human partner.

Since Kismet, there have been several robots that employ similar affect-based architectures and emotion expression systems. We discuss these in the following section alongside other computational HRI systems concerned with expressing emotion more generally, as well as those that tackle the converse problem: the detection of emotions in a social context. Detecting emotions in HRI overlaps significantly with the larger body of work on emotion recognition in human-computer interaction and affective computing. Similarly, the challenge of emotion expression is extensively studied in the virtual character and embodied conversational agent literature. In this survey, we do not cover these two literatures, but instead focus on the work done specifically in the context of HRI.

5.1 Models of Emotion for Social Robots

When a social robot interacts with a human, the benefits of a computational capability to express and perceive nonverbal emotional cues are evident. In contrast, the utility of deeper underlying emotional models is more subtle. Like any control architecture, computational emotion models are designed to govern a robot's behavior, i.e., its response to external stimuli. This can have two potential outcomes, expressive and pragmatic: First, emotion models allow a principled parametrization of socially expressive behaviors. Second, these models can drive decision-making for action planners and regulate other pragmatic behaviors, such as attention.

Many researchers are influenced by and build on Kismet's emotion architecture, which maps emotions on a multi-dimensional space [Breazeal and Scassellati, 1999]. A popular such affective space is along the *arousal-valence-stance* dimensions [Scherer, 2005]. Others [e.g., Saldien et al., 2010] use the circumplex model of emotions [Russell, 2003].

Kim et al. [2005] propose a two-layer architecture with an internal drive system inspired by that of Kismet. The *reactive* layer encodes predefined rules which relate input stimuli to corresponding emotional expressions. The *deliberative* layer models appraisal of stimuli and operates through "action coloring," i.e., the expression of emotions through the modification of other actions [see also: Park et al., 2009]. Hirth et al. [2011] present a framework that models five different functions of emotion, namely the *regulative*, *selective*, *expressive*, *motivational*, and *rating* functions.

Some researchers broaden the scope of affect models for robots by considering slower-changing and longer-term properties of biological emotion systems. This includes models of *mood*, *attitudes* and *personality*. For example, Moshkina and Arkin [2005]'s TAME framework (Traits, Attitudes, Moods and Emotions) models mood as undirected, longer term, lower intensity emotional states. In their model, attitudes are predispositions towards positive or negative emotional states in reaction to certain objects, people, or situations. Similarly, Gockley et al. [2006] present an emotional model with short-term responses, mid-term moods, and longer term attitudes, implemented on a robot receptionist. Ahn and Choi [2007] represented personalities as different parameter values of a linear equation that combines a system of state dynamic equations corresponding to reactive, internal, emotional, and behavior dynamics.

Common to the above-listed cognitive models of emotions is that they are highly sensitive to well-tuned parameters. In other words, to enable a system to react to stimuli in a desired way, the parameters of the system need to be set in a particular and often hand-coded way. These parameters include, for example, the arousal-valence-stance tags associated with different stimuli, the weight that each input has on the underlying motivational drives, and the decay factor of drives and affects. The parameters are also not necessarily stable. Small changes to parameter values may result in vastly different robot behavior.

There is currently no systematic theory or methodology for how these parameters should be tuned to appropriate values or potentially learned from the robot's past interactions. This is a promising opportunity for future research and a necessary one toward the broader use of emotion models.

5.2 Expressing Emotions to Communicate with Others

A large portion of emotion-related work in computational HRI focuses on robots effectively expressing emotions across modalities. The majority of this research focuses on facial expressions and motion generation, with a smaller share using other channels, such as vocalics and color for expressing emotions.

5.2.1 Facial expressions

Facial expressions are a central modality for emotion expression [see also: Ekman and Friesen, 1969], and are therefore heavily used in HRI as well. For example, a long-term deployed robot receptionist expresses emotions on a screen-based avatar through facial movements as part of direction giving dialogs with visitors [Gockley et al., 2006].

A great number of works involve using hand-crafted facial expressions for high degree-of-freedom anthropomorphic faces, often in the context of a new robot face being developed [Lütkebohle et al., 2010, Ahn et al., 2012, Kedzierski et al., 2013]. To determine a more general approach, Bennett and Šabanović [2014] study the minimal set of features for expressing a full range of emotions and find that movement of the eyes, eyebrows, and mouth alone were sufficient. Moving beyond manually specified discrete facial expressions, Shibata et al. [2006] presents a parametric model for generating emotional facial expressions. Their facial features are implemented in the form of LED projections on an anthropomorphic robot.

A challenge for designing robotic facial expressions is that their understanding by humans heavily depends on context and combination with additional communicative channels. Costa et al. [2013] demonstrate that gestures that accompany facial expressions can aid the recognition of the expression. Zhang and Sharkey [2011] show that music with congruent valance can enhance the recognition of a robot's expression.[1] Despite these studies, we did not find works that specifically use context in support of computationally generative facial expression systems.

5.2.2 Affective Motion

Another potential channel for emotional expression is a robot's motion path. Unlike faces, usually purely intended for communicative expression, a robot's motion often has a pragmatic purpose and thus additional constraints. Hence, expressive motion needs to be overlaid over or sequenced with pragmatic motions. An additional challenge is that a robot's motion capabilities are extremely diverse (for example, robotic arms and mobile robots have a vastly different motion space) and are also usually quite different from human motion.

HRI user studies show that emotional state can be legible through posture [Breazeal et al., 2007], whole body locomotions [Kishi et al., 2013], and emotive gait patterns [Karg et al., 2010], several of which have been used to generate expressive robot motion paths.

For example, Bretan et al. [2015] present a human-inspired emotion expression system for a non-humanoid desktop robot. The framework manipulates eight movement parameters (including posture, head

[1] Context is also a confounding factor in the evaluation of an emotionally expressive system. Any design of a computational system to express emotions carries with it the challenge of how to evaluate its success. In a typical setup, researchers ask humans to judge what emotion is being expressed by the robot, but this approach is context sensitive. Whether the expression is experienced in a contextual vacuum, alongside other cues, or within an ongoing interaction will bias how a person judges an expression. A second factor is dynamics: Whether a person views a static picture, a dynamic video, or a situated interaction of the expression will bias their perception of the expression [see: Bretan et al., 2015]. These and other confounding factors should be considered when discussing the evaluation of any emotionally expressive computational system.

activation, volatility, and exaggeration) based on the circumplex model of emotions. They then overlay these parameters over the robot's pragmatic movement.

Sharma et al. [2013] develop guidelines to author aerial robot motions to elicit desired affective responses. Their work is based on the Laban Effort System (LES) by Laban and Ullmann [1971], which is increasingly adopted in HRI research for modeling emotion expression through motion. The LES modulates movement through four parameters: Space, Time, Weight, and Flow. Other work using the LES includes Knight and Simmons [2014], who aim to automatically produce Laban effort motion paths for point robots, and Masuda and Kato [2010] who present a method for adding a target emotion to arbitrary motions of a humanlike robot.

5.2.3 Expressing Emotion through Other Channels

A robot has access to additional nonverbal channels to express emotional state. To support this claim, a study by Niculescu et al. [2013] demonstrate that different robot personalities could be attained by manipulating voice pitch. Johnson et al. [2013] demonstrate the successful use of non-anthropomorphic colored LEDs. They manipulate the color, intensity, frequency, sharpness, and orientation of movement of LEDs around a robot's eyes to express emotions. Despite the above studies, we did not find generative computational systems that make use of these channels for robot emotion expression.

5.3 Recognizing Emotions in a Human Partner

Recognizing a human's emotional state in an interaction with a robot has several uses. In the context of collaboration, it can allow a robot to adapt its behavior to fit the person's state and preferences. In a learning interaction, it can serve as a natural channel for human feedback about the robot's execution of newly learned actions. As mentioned above, the work in computational recognition of human emotion is part of the broader topic of affective computing. We discuss only the work related to perceiving human emotions within a social robotics setting.

Many research projects focus on facial expressions and facial feature extraction from vision during human-robot interaction [e.g., Cid et al., 2013]. For example, Lang et al. [2013] enable automatic recognition of facial communicative signals in the context of an object teaching scenario, comparing a static approach to one that considers temporal dynamics. In other work, the real-time emotion recognition system (RTERS) localizes faces and extracts their features in a sequence of images. It then codes facial expressions into one of seven different emotional states: happiness, sadness, fear, disgust, anger, surprise, and neutrality [Alazrai and Lee, 2012]. In Boucenna et al. [2014], a robot learns to recognize the facial expressions of the human partner on-line if they imitate the robot's prototypical facial expressions.

Within the visual modality, emotions can be determined by features other than facial expressions. Sanghvi et al. [2011] analyze features from videos extracting postures and body motion to detect engagement of children playing chess with a robot. McColl and Nejat [2012] classify real-time body language using a RGB-D sensor. Their system determines a person's affect in terms of their accessibility towards a robot during one-on-one interactions. Beyond postures, Venture et al. [2014] look at body motion and demonstrate a system that recognizes affect from gait. They conclude that it is possible to discriminate affects from gait data significantly better than chance, using only the lower torso movement and the trunk and head inclination.

Finally, there is a body of work in detecting affect from physiological signals. Leite et al. [2013] describe a robot that plays chess with children while monitoring their electrodermal variations as an indicator of their affective state. In Liu et al. [2008], experimental results with six children on the autism spectrum show that a robot can automatically predict individual anxiety and liking level in real time with 80% accuracy based on physiological signals of electrocardiography (ECG) and electromyography (EMG) during the interaction. Finally, Kulic and Croft [2007] present an HMM-based affective state classifier for estimating affective state from physiological data during human-robot interaction. Their sensors include skin conductance, heart rate, and facial muscular movement detectors using EMG.

To summarize, our survey found work in computational HRI that is related to emotions in three areas: underlying models of emotions and affect for modulating behavior; systems for expressing and communicating emotions, mainly through facial expressions; and systems for detecting emotions in humans, in the context of interacting with a robot.

In the emotion expression space, much of the work focuses on human-inspired robot faces, whether projected on a screen or implemented using mechanical actuators. However, we know from popular culture, such as science fiction and animation, that non-anthropomorphic robots can also be emotionally expressive using only audio and motion cues. Until very recently, there was little work in these areas compared to the large amount of research on human-like facial expressions. In addition, this question is likely to be more pertinent as low-cost robotic devices are developed for home use. These domestic robots will have a low-degree-of-freedom design and will need to express emotional state under these constraints.

Furthermore, all three thrusts (models, expression, and recognition) have been mostly studied in isolation from each other and outside the context of entire interactions or task collaborations. However, human subject studies show that context greatly impacts people's emotional state and emotion perception. This suggests that an integrative approach to computational emotion research in HRI could be a fertile ground for future work. We also identify a lack of literature on the topic of machine learning aimed to automatically determine the parameters of emotional models for social robots. Finally, there is an open area of study in computational models of emotions for HRI in real-world environments.

6

Understanding Human Intentions

In the rest of this survey we focus on the high-level social capacities and behaviors that are predicated on the foundations we have discussed thus far. The first of these is the capacity of a robot to make inferences about human intentions and to communicate its own intentions in an interaction.

Humans have a natural tendency to interpret the behaviors of others as *intentional*, *goal-directed* actions. Studies of infants show that they are able to segment complex actions into units corresponding to the initiation and completion of intentional action [Baldwin et al., 2001]. They show surprise when someone executes actions that are inefficient in achieving goals [György et al., 1995], and are more likely to imitate actions they perceive as intentional than those they perceive as accidental [Carpenter et al., 1998, Tanya et al., 2005]. This indicates that the ability to perceive and reason about intentional action is central to the way people interact with and around each other from a very early age.

Philosopher Daniel Dennett calls this ability to infer intentions the “Intentional Stance" [Dennett, 1989]: As social animals, humans developed the ability to reason about mental states—beliefs, desires, and intentions—in order to predict the actions of other humans and animals. Humans apply this reasoning strategy to anything that produces self-motivated action that cannot be described by physics. For example, a stone rolling down a hill is doing so because of physics, whereas a person grabbing an apple is doing so because they want to possess (and perhaps eat) the apple. Interpreting the behavior of others as goal-directed enables an observer to infer additional meaning about the behavior of others, and plays a crucial role in collaborative and social behavior [Tomasello et al., 2004]. Hence, the construct of intentional action is also an important aspect of social robotics. In this section we survey the research around representing, understanding and reasoning about the intentions of a human partner, as well as research concerned with enabling robots to communicate their own intentions.

6.1 Toward a Theory of Mind: Cognitive Frameworks for Intention Parsing

The ability to decipher a mental state of another person is called "folk psychology" or denoted as having a Theory of Mind (ToM) capacity. While some aspects of ToM are present from a very young age (e.g., gaze following, joint attention, and social referencing), it is widely accepted that full ToM capability develops progressively until the age of five [Gopnik et al., 2001]. Over the years, several cognitive architectures for HRI were inspired by the notion of ToM and implemented aspects of it to allow robots to parse intentional behavior.

For example, Trafton et al. [2013] describe the ACT-R/E cognitive architecture for HRI. The motivation is for robots to build spatially embodied models of human cognition in order to understand how and why people think the way they do. Robots can use this kind of perspective taking—and reasoning about people's reasoning—to predict what a person will do in different situations, e.g., that a person may forget something and may need to be reminded, or that a person cannot see something that the robot sees. Kennedy et al. [2009] gives three examples of specific social situations in which ACT-R/E is able to represent and reason about the following: spatial perspective taking, action perspective taking in teamwork reasoning tasks, and dominant-submissive reasoning in a social task. In all three cases the system uses its own model of the skill as a simulation of the human partner in order to infer things about the given situation.

On that note, Pandey and Alami [2010b] compute “mightability maps”, enabling a robot to reason about what a human collaborator might see or reach in a shared workspace. These are computed using the human’s position, posture, and visual perspective-taking. In the context of shared Urban Search and Rescue (USAR), Talamadupula et al. [2014] use probabilistic reasoning about both the human’s and robot’s beliefs to decide between alternative plans with ambiguous commands.

Some researchers in computational HRI use “Simulation Theory” as a theoretical construct to underpin ToM. The theory is grounded in the idea that humans use motor and perceptive resonance, i.e., simulating a collaborator with one’s own motor and sensory-perceptive cortex to enable joint activities [see: e.g., Sebanz et al., 2006].

One such example is work by Breazeal et al. [2009], who propose an *embodied cognition* architecture that enables a robot to take the perspective of a human collaborator. The architecture is inspired by human ToM models, in particular the reuse of the robot’s own action system to simulate the likely intentions and goals of the human collaborator. The robot simultaneously holds its own beliefs and the beliefs of the human by simulating the human’s sensory perception through its sensory and perceptual system, and additionally using information about the human’s sensory perspective. This architecture was implemented on a humanoid robot in collaborative and learning interactions.

Gray and Breazeal [2014] also present a self-as-simulator architecture for mental state manipulation through physical action. The robot models how a human’s mental states are updated through their visual perception of the world around them. This modeling, combined with geometrically detailed perspective correct simulations of the immediate future, allows the robot to choose actions which influence the human’s mental states through their visual perception. The system is demonstrated in a competitive game scenario, where the robot attempts to manipulate the mental states of an individual in order to win. Milliez et al. [2014] show a similar result by maintaining a human-compatible representation of the world and people and objects, allowing for appropriate reasoning about intentions. Their system, called SPARK (SPAtial Reasoning and Knowledge), can represent objects in the world, generate relative and symbolic information about these objects for the purpose of communication, and reason about humans in the environment in terms of their visual perspective and beliefs about these objects.

Others have shown that a full-blown cognitive architecture may not be necessary to still achieve some of the benefits of self-as-simulator for intent recognition. In Kelley et al. [2008], the robot first learns activities by building a Hidden Markov Model activity model. Then it uses the same set of models to recognize an intent as seen by a human in the environment. The key to this is the level of representation used. The robot's motion and human's motion are both seen as a function of the position and orientation of the actor (robot or human) with respect to another person in the environment. Butterfield et al. [2009] propose that Markov Random Fields (MRFs) are a better probabilistic mathematical model for incorporating the internal states of other agents into robot decision making. They propose that ToM capabilities are essentially the estimation of latent variables based on the history of perceived observations and present theoretical models that capture the experimental findings from Theory of Mind studies in developmental psychology.

6.2 Parsing Human Attention

Full ToM and the ability to infer a variety of mental states is more than most social robots can achieve today. As a stepping stone, much work has gone into endowing robots with the ability to achieve one particular mental state inference: inferring what the human is attending to in order to establish joint attention between the human and the robot. This is clearly the precursor to much of social interaction; a robot has to infer, represent, and reason about what aspects of the environment an interaction is *about*. Attention parsing has been a challenge for the computational HRI research community for many years [see: e.g., Scassellati, 2001]. In this section we cover only a few recent examples.

Nagai [2005] presents a supervised learning approach, building a model that can reflexively achieve joint attention. It uses the coordination of two neural networks, one modeling edge features, the other modeling optical flow. The learned models detect a gaze direction from a camera zoomed on the face, which is then used to select an object of attention from a camera that has a wider view of the person and the workspace. Results indicate that both modalities together achieve better performance than either one alone.

Huang and Thomaz [2011] outline a tripartite model of joint attention capabilities for social robots: responding to joint attention, initiating joint attention, and ensuring joint attention. In a user 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 ensuring joint attention behavior is judged as having better performance in an interactive task and is perceived as a natural behavior.

Joint attention is especially important for shared manipulation of an object. Grigore et al. [2013] study a handover task; their work compares a model that uses an HMM based on only physical features of the action versus one that uses information about the human's engagement in an interaction: eye gaze and head gaze orientation as a sign of a human's focus of attention and engagement in a task. Their findings indicate that the models incorporating attention signals are significantly more successful for handing over an object.

Broquère et al. [2014] present an attentional controller for human-robot interaction. Their system first computes a cost-map for each point in space based on the human's visibility and comfort, as well as the robot's end effector and object positions. These cost maps are used to adapt the sensor processing frequency of that region, as well as the monitoring frequency parameters of various behavioral primitives. An executive then takes into account the various primitives' states to both switch between them, and adjust the parameters sent to a planner which is executing the chosen primitive.

6.3 Understanding Intentional Action for Prediction

Deciphering the intent of a human action goes beyond recognition of the current activity, and includes inferring the goal a person is trying to achieve and predicting their future actions. Many works in this area focus on computational models that allow for forward prediction of the human's activity, letting the robot anticipate their action before it happens.

Traditional approaches to activity recognition require seeing the entire activity before a classification can be made. In contrast, Ryoo et al. [2015] detect onset signatures of activities enabling the robot to detect them before completion. This is an example of incremental recognition, which highly relevant to the problem robots face in an HRI setting, i.e., making real-time predictions of human partners. Relatedly, Iengo et al. [2014] use an HMM-based approach that can be trained online with few samples and can cope with intra-user variability during the gesture execution. Models are employed within a continuous recognition process that provides the probability of each gesture at each step.

Hoare and Parker [2010] use Conditional Random Fields (CRFs) to determine the human's intended goal, showing the effects of using different task-related features to improve accuracy and the time to the correct classification. They show that CRFs work well for classifying the goal of a human in a box pushing domain where the human can select one of three tasks, and that the correct classification can successfully happen online before the task has completed. The approach in Bascetta et al. [2011] similarly aims to estimate the intention of a human in a confined space based on their motion trajectories as viewed from an overhead camera. Based on offline observations of behaviors in the space, HMMs are trained to model typical activities. These models are then used to predict the most likely final locations of the human given an initial motion trajectory, allowing a robot in the space to anticipate where the human will be and plan its actions to ensure safety.

Nyga et al. [2011] propose a system for clustering and semantically annotating trajectories observed in human manipulation activities. The system learns models of human motions in the context of complete activities and is able to robustly cluster noisy trajectory data obtained from real-world observations. The learned models can then serve as predictive models for human motions.

Koppula et al. [2013] present an approach for jointly labeling human sub-activities and object affordances in order to obtain a descriptive labeling of the activities being performed in RGB-D videos. They formulate this problem as a Markov Random Field and learn the parameters of the model using a structural Support Vector Machine formulation. The model also handles segmentation, computing multiple segmentations and treating labels over these segments as latent variables. Koppula and Saxena [2016] build on this work to allow a robot to anticipate a human partner's future activities. Using a CRF approach to model the seen activity of the human, and given activity and affordance labels, they sample from the set of most likely target positions of the current objects in the activity to generate future trajectories of the human's motion.

Wang et al. [2013] introduce the Intention-Driven Dynamics Model (IDDM) to probabilistically model the generative process of intentional actions with an approach that learns a model of the intentional actions based on Gaussian Processes, and then uses a Bayesian approach to infer intentions from observed movements. The IDDM simultaneously finds a latent state representation of noisy and high-dimensional observations and models the intention-driven dynamics in the latent states. They show an efficient online algorithm that allows for real-time intention inference.

Najmaei and Kermani [2010] use a neural network to predict the next three human positions in a collaborative workspace based on the previous five positions. They combine their prediction with an impedance based controller avoiding the human's predicted position, primarily for safety applications. Mainprice and Berenson [2013] use a swept volume, based on predictions from a Gaussian Mixture Model (GMM) for the human pose, to define a probable space for a human to operate in. This is then used for safe trajectory planning for close-collaboration robot arm motions.

6.4 Communicating Intent

Finally, we turn our attention to robots that *generate* actions that will be perceived as intentional by a human collaborator. This includes work aiming to produce robot behavior and motion trajectories that are more natural and human-like. The field of character animation has a long history of generating human-like behavior and communicating intent with non-human characters. In their book "The Illusion of Life", Thomas and Johnston [1981] operationalize the artistic process employed by classic 2D animations. Several years later, Lasseter [1987] wrote a seminal paper describing how these insights can translate to 3D animation. Several HRI researchers have worked toward transferring these insights from animation principles to the generation of lifelike robot motion.

Takayama et al. [2011] use the animation techniques of anticipation and reaction to create more readable robot motions. In this work, behaviors are created by an expert animator, showing forethought and goal-oriented reactions to task outcomes. Similarly, Ribeiro and Paiva [2012] present an overview of animation principles and examples of manually creating facial expressions based on these principles.

Also inspired by animation, Gielniak and Thomaz [2012] show that exaggerated motion, algorithmically created through a PCA analysis of the torque trajectory, is perceptually different to a human observer. They also find that this has an impact on directing people's attention to salient aspects of the interaction in a storytelling task, as measured through eye tracking and their ability to recall task elements. Szafir et al. [2014] present two studies examining the effects of modifying the trajectories and velocities of flight primitives for unmanned arial vehicles (UAVs) based on natural motion principles: arcing, ease in/out, and anticipation. Their studies show these manipulations to increase people's ability to infer intent, and the motion was perceived as more natural. Importantly, these last two examples represent autonomous algorithms for generating lifelike motion based on animation principles, and does not require an animator to manually create the motion.

Gielniak and Thomaz [2011b] show that spatiotemporal correspondence (STC) of actuators in a kinematic chain is an important component of generating humanlike motion. Optimizing motion to increase STC causes it to be recognized as more humanlike and more accurately identified as the intended motion. Somewhat in contrast to the above STC results, Riek et al. [2010] perform an experiment studying the smoothness and precision of a humanoid robot's gestures, finding that people are faster to respond to abrupt "machine-like" gestures compared to smooth ones.

Another factor in humanlike motion is variability. Gielniak et al. [2013] add an additional step after the STC optimization mentioned above. With the aim of avoiding repetitive motion, the algorithm adds variance through exploiting redundant and underutilized spaces of the input motion, which creates multiple motions from a single input. As a final step, their algorithm ensures the robot can satisfy task constraints while maintaining the human-like qualities of STC and variance. Minato and Ishiguro [2008] also look at motion diversity for android robots. In the context of a robot reaching toward a known person or a stranger, they modeled the variance in the person's motion and show that reproducing these different motions changes the way people perceive an android. This shows that humanlike motion has this diversity, and motion itself can communicate social context.

In the context of a robot identifying, grasping, and placing objects, Beetz et al. [2010] first suggest the notion of legibility of motion, to help the human predict the robot's trajectory and goal position. They propose to use stereotypical movements based on human placement of objects to introduce legibility. Gielniak and Thomaz [2011a] have a similar approach, with the goal of allowing a human viewer to understand the intent of an arm gesture as quickly as possible. Their algorithm generates an anticipatory motion from a given input trajectory by extracting the intent symbol, which is assumed to be captured in the canonical pose of the hand throughout the gesture. The algorithm uses a motion graph to move the production of this symbol earlier in the motion.

In addition to generating motion plans that take a human's position, pose, and eye gaze into account as optimization criteria, Dragan et al. [2013] have introduced an approach where a robot can make its intended reaching motion clear by optimizing for legibility. They distinguish this from predictability, which optimizes for what an observer would expect given a single goal, whereas legibility optimizes the trajectory such that a human viewer can identify the goal target as early as possible in the face of many possible targets. A similar method also enables solving for legibility when determining the hand position for a robot pointing to an object [Holladay et al., 2014].

While the above-mentioned work focuses on manipulator motions, Szafir et al. [2015] explore the design space of explicit robot communication of flight intentions to nearby viewers of UAV robots. Taking biological flight and airplane flight as inspiration, they develop a set of signaling mechanisms for visually communicating directionality. Kato et al. [2015] also look at the intentionality of motion paths, but in the context of a mobile robot in a shopping mall. Their goal is to model polite approaching behavior. An analysis of human shop attendants revealed their attention to intentions of nearby visitors. The author's modeled these behaviors, produced a robot implementation, and tested them in a shopping mall.

The key idea for all of the above is that the robot's intention be transparent, letting a human partner readily infer the intended target or goal of its action. Related to this, several have studied the opposite of transparency: deception. For example, Dragan et al. [2015] use the legibility optimization introduced in earlier work to do the opposite optimization to hide the robot's intent such that an onlooker cannot tell until the last minute what object the robot will grab.

To conclude, this section surveyed the construct of intentional action in computational HRI. One side of this involves representing, understanding and reasoning about the intentions of human collaborators, a capacity called Theory of Mind (ToM). Simulation Theory is a popular approach to cognitive architectures for understanding intentions via ToM. A simplified version of ToM is to only infer attention of the human partner and the joint understanding of what object an interaction is about. Simplified further, many approaches to intentional understanding have no basis in ToM and instead focus on representing activity with probabilistic models that allow for early prediction of action completion. This subfield is still in its infancy, and many opportunities remain in understanding human intentional action. Greater use of context and multiple modalities stand out as important avenues yet to be explored in depth. Additionally many approaches have focused on low-level movement and action prediction. Moving beyond, toward higher level activities could allow for more useful real-world predictions of intent. Instead of only detecting which object will be manipulated several milliseconds or seconds early, the robot could add predicting the next five most likely object manipulation actions based on the task and context.

The other side of intentional action is having robots generate actions that will be perceived as intentional by a human collaborator. Our survey has found computational approaches to this involving optimization techniques inspired by character animation principles, and optimization for legibility. The goal of these works is making the intent of the robot as transparent as possible to the human, as early as possible in the interaction. Future work in the domain could turn its attention to the many additional animation principles that have not been algorithmically modeled in computational HRI. Additionally, there is a need to move beyond object-directed reaching motions when working toward legibility, generalizing to other classes of intentional action as well as to task-level longer-horizon expression of intent. Multimodal communication of intent is unexplored for the most part and may also play a larger role in future research.

7

Human-Robot Collaboration

Building on the capacity to infer, reason about, and generate intentional action, a major research thrust in computational HRI focuses on developing and studying robotic systems that engage in human-robot collaboration. Extending the traditional planning and execution literature, researchers propose cognitive frameworks and other computational architectures enabling and supporting teamwork. Many of these are inspired at least in part by human-human teamwork. More recently, researchers have also started to address questions of collaborative timing and fluency, as well as the question of how to plan a single action motion plan in the context of a human collaborator. We find that much of the literature gives special attention to two particular collaborative tasks: handovers of objects and collaborative manipulation. These will be discussed at the end of this section.

7.1 Planning and Execution Frameworks for Collaborative Activities

Making plans for action is at the core of autonomous robotics. It is often framed as a problem of generating the optimal action for a given, possibly non-deterministic and partially-observed environment. Planning for a joint activity with a human necessitates a reformulation of the notion of "environment", as the human is not only dynamic and non-deterministic, but also perceives the robot, adapts their own plans and actions, and usually has a common goal with the robot. The robot needs to plan in a way that coordinates and meshes with the human collaborator's activity.

Hoffman and Breazeal [2004] address this issue by building on the notions of Shared Cooperative Activities [Bratman, 1992], Joint Intention Theory [Cohen and Levesque, 1991], and Dialog Theory [Clark, 1996]. They identify joint intentions, meshed subplans, mutual belief, common ground, joint closure, and mutual support as foundations for human-robot teamwork. Based on this, they present a hierarchical goal-oriented task execution system integrating pragmatic action with human verbal and nonverbal communication, as well as robot nonverbal communication supporting the shared activity requirements. The system also allows for dynamic agent allocation for each task based on the information inferred from the verbal and nonverbal channels. In related work, Lenz et al. [2008] present a human-robot collaborative system which includes explicit modeling of joint attention, shared tasks, and action coordination through communication. This enables multi-modal shared factory-like activities. It primarily uses task structure to anticipate the human's next action.

Schrempf et al. [2005] suggest a framework for human-robot collaboration using human intention recognition with Dynamic Bayesian Networks (DBN). A planner uses minimum entropy to choose between competing human intentions and executes a hand-coded action from a fixed set. Similarly, Schmid et al. [2007] estimate the human's intention as a probability density function over the possible intentions. They also use an entropy cost function to proactively execute an action when there is a small number of competing intentions.

Shah et al. [2011] present a goal-oriented task-level controller, originally developed for multi-robot synchronization, for use in a human-robot collaboration setting. The controller works by computing a compact representation of the shared plan offline, and by then modifying the timing and goal-state in real-time during human-robot collaboration. This allows for flexible teamwork through dynamic allocation of sub-tasks, as well as for communication about shared and individual goals.

One aspect of teamwork is the allocation of actions in a shared plan, and whether actions are mutually exclusive or can be completed by either team member. Nardi and Iocchi [2014] deal with “social plans” — plans that have actions to be executed by the robot and actions to be executed by a human collaborator. They model them as multi-agent Petri Net Planners (PNP) and suggest a method by which the shared PNP is converted to a robot-only PNP. The method includes inserting a preparation action (e.g., approaching a human), a communication action (e.g., asking for help), and a perception action (e.g., perceiving the outcome of the human action). They implement their system in an office navigation scenario where the robot needs humans to press elevator buttons for it.

To support action sharing in joint activities, Nikolaidis and Shah [2013] model cross-training—performing the teammate’s role instead of the robot’s own—to improve human-robot coordination. The robot’s policy is updated using the human’s actions under the assumption that these would be what the human expects the robot to do. In a different line of research, Nikolaidis et al. [2015] use a two-phase approach to fit a robot’s collaborative policy to a human collaborator: First, the human activity is clustered into collaborative “styles”, resulting in a Mixed Observability Markov Decision Process (MOMDP) policy with the style as the hidden variable. During the second phase, the robot infers the human style and uses the appropriate policy, taking into account the uncertainty about the style inference.

In Ben Amor et al. [2014], joint physical activities are modeled using “Interaction Primitives”, extending Dynamic Movement Primitives (DMP) previously used for single robot learning. First, reference trajectories for the human’s and the the robot’s DMP are collected from human-human demonstration. Then, Dynamic Time Warping (DTW) is used to estimate the human’s current phase in their DMP. The robot’s DMP parameters are then predicted based on the predicted human's DMP parameters and the current phase estimate. Finally, in the joint music-playing domain, Hoffman and Weinberg [2011] suggest achieving simultaneous yet reactive shared activities for an autonomous robotic jazz-improvising robot. They propose three principles to support this goal: anticipatory action, simultaneous learning and acting, and embodied opportunism, and implement each in a separate robotic improvisation module.

7.2 Timing and Fluency

Extending beyond action selection and discrete meshing of single actions withing a task, the computational HRI literature has recently started to address questions of action timing. Within that context, *fluency* is often described as a goal or metric of joint action between a human and a robot [Hoffman, 2013]. Intuitively, a well performing team displays fluidly seamless interaction, but it is in practice it is difficult to measure and optimize for fluency. Hoffman and Breazeal [2007] describe an anticipatory action system based on a two-agent Markov Decision Process (MDP) representation. They suggest a policy learner minimizing the expected cost of acting using a Bayesian model of human action sequences. They find that even when task efficiency was not significantly affected, people's sense of the collaboration was. The authors suggest the possibility that people are sensitive to an alternative quality they call "fluency", and propose metrics to evaluate this quality. In a continuation of that work, Hoffman and Breazeal [2010] present a connectionist embodied cognition framework to achieve anticipation and fluency in a continuous human-robot shared task. Modeling the human phenomenon of perceptual simulation, sensory channels feed bottom-up activation nodes, which are simultaneously affected by the robot's prediction of human action in a top-down manner. Hebbian reinforcement triggers cross-modal activation leading to higher human-robot efficiency and fluency.

More recently, researchers extended probabilistic reasoning frameworks to include information about timing. Hawkins et al. [2013] offer a graph representation modeling the timing of the human's actions as random variables, and also taking into consideration action preconditions and sensor errors. They present a planner using a cost-minimization criterion. While this work is only able to model predefined (and known) sequences of actions, Hawkins et al. [2014] extend this work to hierarchical representations of tasks as AND-OR trees, using a similar random variable representation for action duration. Kwon and Suh [2012] extend Bayesian Networks to include a random variable for the time of an event, which is useful for modeling the uncertainty about the human relative to both action selection and to temporal occurrence ("whether" and "when"). This is used by the robot to generate anticipatory action to minimize both human and robot waiting times.

Examining the timing structure underlying turntaking in a multimodal interaction from a human-robot study, Chao et al. [2011] identify the minimum necessary information (MNI) as a predictor of when to act. Based on this insight, Chao and Thomaz [2012] address the problem of timing in multimodal human-robot interactions using Timed Petri Nets (TPN). This structure enables the modeling of action durations, dependencies, and delays. They use this model to overcome action atomicity and enable the interruption necessary for fluent collaborative activities. In particular, they enable the robot to detect human intent to act and interrupt previously started actions. The CADENCE system [Chao and Thomaz, 2013] complements the above work by integrating elements into the TPN architecture to specifically model the management of the collaborative "common floor". The robot is able to reason about seizing, holding, yielding, and auditing the floor, which can belong to the human, belong to the robot, or be in conflict or lapsed. It can then use functional actions or backchannel communicative acts to regulate the common floor across multiple modalities, including speech, gaze, and gestures.

Music is a particularly time-sensitive domain, and theorists have explained the appeal of music to humans as indicative of the importance of human cognition about temporal patterns [Minsky, 1982]. Several robots have been designed to move to a musical beat. Keepon, a small non-humanoid robot synchronizing its movements to the dominant beat in the environment, has been shown to have an effect on children's interactivity with the robot [Michalowski et al., 2007]. Also using

Keepon, Avrunin et al. [2011] showed that perceptions of the lifelikeness of the robot and the quality of the dance can be manipulated by small variations in the robot's movement timing. Hoffman [2012] presents an autonomous smartphone-based robot capable of generating dynamic choreographies based on a song's rhythmic patterns and genre, and this research shows that the robot's synchronized movement with music has effects on people's opinion of the music and of the robot's character traits [Hoffman and Vanunu, 2013].

7.3 Human-aware Motion Planning

The frameworks discussed thus far focus on task-level action meshing between a human and robot agent working together. Collaborative systems can also modulate each primitive action to support the joint activity. Traditional motion planning is usually framed as a search or optimization problem, and the criteria that deem one trajectory or motion path more optimal than another are based on task or workspace constraints. In a human-robot collaboration, however, robots generate actions that are viewed by a human partner. As a result, the design of these actions should take into account people's propensity to infer intentions and goals, as well as their safety and overall usability. This section complements Section 6.4, which deals specifically with generating intentional action.

Based on the insight of an observer, Sisbot et al. [2010] present an integrated motion synthesis framework that is especially designed for a robot that interacts with humans, dealing with perspective taking, human-aware manipulation planning, and soft trajectory planning. The resulting system generates robot motions taking into account a human's safety, their vision field and perspective, their kinematics, and their posture comfort along with task constraints. In a similar spirit, but focused only on safety, Ding et al. [2011] show a method based on HMMs to predict the region of the workspace that is possibly occupied by a human within a prediction horizon; they then use this prediction to construct safety constraints for an industrial robot.

In addition to taking the social context into account at planning time, many works stress the importance that a robot's motion planning be able to handle interruption. Kondo et al. [2013] use gestures obtained by motion capture of five people, parameterized with different target poses or directions. Their experiments show that motion parameterization increases the number of people that voluntarily interact with the robot, and interruptibility increases the duration of these interactions. In Xu and Dudek [2012], a robot's reputation (defined as the human's trust in the robot) is modeled and updated based on the amount of overrrides observed by a human supervisor. The robot uses this signal to adapt its planner policy via learning.

7.4 Object Handover Actions

One of the most studied human-robot collaborative tasks is handovers, including both handing objects to a robot and receiving objects from a robot. In their seminal work, Edsinger and Kemp [2007] demonstrate a full-circle system approach for a robot taking objects from humans. The robot seeks out a human in the workspace using visual sensing, reaches toward the person, then detects the human's hand velocity, using vision, and subsequently lowers its own arm stiffness in order to accept the object until a successful grasp is detected in the robot's hand sensors.

Strabala et al. [2013] provide a good overview of the human-robot handover literature. In their own work, they observe human-human handovers, identifying three activities: carrying, coordinating, and transfer. They also use learning via feature selection to automatically predict handovers with close to 90% prediction rates. In related work, Cakmak et al. [2011] evaluate spatial and temporal contrast in human-robot handovers to improve the fluency of the action and find that temporal contrast reduces human waiting times, but that spatial contrast is not effective.

Sisbot and Alami [2012] describe a handover planner that takes into account human safety and comfort as well as the legibility of the robot handover trajectory. The planner first determines the object's handover point by taking into consideration a human safety, visibility, and comfort map. This then determines the object's trajectory modeled as a free-flying object. Finally the robot's trajectory also takes into consideration the legibility of the handover. The authors suggest planners based on workspace grid search, as well as on Rapidly-exploring random trees (RRT) on the robot configuration space [see also: Mainprice et al., 2011]. In a similar vein, Williams and Breazeal [2012] take into account object, robot reach, human grasping, and social considerations when planning how to grasp an object for handover to a human.

Chao et al. [2013] use a people-tracker instrumented field observation to build a model of leaflet handover in a public space, looking specifically at the relationship between gaze, arm extension, and approach. They implement their findings to build a robot handover and gesture controller deployed on a small humanoid robot, showing an increase in accepted leaflets by passers-by.

Chan et al. [2013] analyze the grip forces and load forces in human-human handovers and characterize the various stages of handovers by the grip-to-load force ratios in giver and receiver. They use their findings to design a robot-to-human handover controller mimicking the discovered behaviors and implement it on a humanoid robot. Subsequent human-subject studies uncover controller parameters for the most preferred handover behaviors with users.

A number of handover works use human data to design their controllers. Yamane et al. [2013] build a motion database collected from human handovers for a data-driven method for motions involved in receiving objects. The database holds the trajectories as a hierarchical search tree. During data collection, the correspondence between the passer and the receiver tree nodes is recorded and then used to generate the receiver trajectory based on the observed human passing patterns. Huber et al. [2009] also collect data from human handovers across a shared table, and find that an axis-decoupled minimum jerk trajectory (MJT) models the handovers better than a regular MJT, which in turn they found to be more effective than a trapezoid trajectory in a previous study. They implemented this control model on a robot handing over blocks to human collaborators. Studying handover timing, Huang et al. [2015] analyze depth sensor data of two people engaging in a handover task. They identify two strategies used by the person initiating the handover when the receiver is not ready to receive the object: waiting and slowing down. They implement this insight in a robot planner that uses these strategies when it detects delay in the human receiver.

Combining handovers and proxemics, Mainprice et al. [2012] model the human's potential movement toward a navigating robot when taking into account the optimal handover position. The cost function includes the human's eagerness to complete the handover as fast as possible, their comfort, and their mobility.

Finally, Aleotti et al. [2014] take an object's grasping affordances into consideration when presenting it to a human for handovers. Using a point-cloud based recognition on eye-in-hand laser scanner data, the robot selects the most appropriate grasp leaving the human-affordable region of the object free. Then a human detector and trajectory planner are used to execute the handover.

7.5 Collaborative Manipulation

Handovers are an instance of a larger body of work that looks at physical contact in human-robot interaction, sometimes dubbed "physical HRI". This area of research also includes the collaborative manipulation of objects. In a large body of control literature, collaborative manipulation is modeled as a dynamics and controls problem, which we do not cover in this paper. However, in some cases the collaborative manipulation challenge relates to the social, intentional, and nonverbal aspects of the activity. In those cases, much of the attention has been paid to the various roles the human and the robot can take on in the joint activity.

For example, Evrard and Kheddar [2009] suggest a homotopy mapping between a controller representing a leader to a controller representing a follower for collaborative manipulation. This enables smooth interpolation between the two control modes. In contrast, Li et al. [2015] propose a game-theoretical approach to dynamically switching the robot's role in collaborative manipulation from a leader to a follower role. They model the shared manipulation as a two-agent game, and use real-time force input from the human to determine the amount of shared control exerted by the robot.

Mörtl et al. [2012] examine role allocation in collaborative manipulation by analyzing human-robot cooperative carrying of a load. They compare three role allocation strategies: static role allocation, continuous adjustment of roles based on the human's haptic expression of acceleration and deceleration, and discrete adjustment of roles. In their analysis of effort sharing, static role allocation means a feed-forward calculation of the applied forces, similar to zero force control. In dynamic allocation, the perceived human feedback is taken into account to generate robot effort in the same direction. This adjustment happens either continuously or after a constant time interval. They demonstrate their system on a robot co-carrying a heavy table with a human collaborator.

Peternel et al. [2014] suggest a tutoring-based method to teach a robot to collaborate on a human-robot joint cross-sawing task, which requires rapidly alternating leading and following with simultaneous motion and compliance adaptation. A human tutors the robot first by teleoperating the robot in collaboration with another human. The human's control is measured visually through force sensors and using electromyography (EMG). After learning, the robot uses adaptive frequency phase oscillators for periodicity and Dynamical Motion Primitives to encode the task execution.

Unrelated to leader-follower roles, Medina et al. [2015] improve on anticipatory control of a robot moving an object jointly with a human by modeling the uncertainty in prediction determined by previously experienced disagreements between the robot's prediction and the human's actual behavior.

In summary, the human-robot collaboration literature makes up a large portion of computational HRI. In many ways, large swaths of the other sections in this paper could be reasonably subordinated to the overarching goal of human-robot collaboration. In this section, we have surveyed works specifically dealing with collaboration, including research concerned with collaborative planning and execution frameworks, enabling robots to mesh their actions with those of a human, and addressing questions of timing and fluency in collaborative scenarios. To a lesser extent, the literature also addresses the characteristics of motion trajectories as part of a collaboration with a human in the loop. We found two widely tackled applications areas of this topic: the handover of an object between a human and a robot and the collaborative manipulation of objects.

While collaboration is a major research area, the application areas and deployment contexts for these systems have been concentrated on a few canonical problems. Most of the collaborative frameworks have not been put to the test in any real-world environment or studied in a long-term setting beyond the lab. Tasks have been simplified versions of the intended scenario. Implementing a collaborative planner in a complex realistic setting would be an apt grand challenge for the human-robot collaboration community. Furthermore, studying additional micro-collaborations beyond handovers and shared manipulation is a wide open research area. Finally, collaborative systems are mostly envisioned for physical tasks (e.g., assembly and manufacturing). Extending the application areas for new kinds of collaborative scenarios, in particular those supporting the service and creative sectors, could help generalize the lessons learned from the traditional scenarios.

8

Social Robot Navigation

Robot navigation is traditionally framed as a spatial planning problem. In order to appropriately plan a path through an environment the robot needs to incorporate dynamic perception of the environment and keep a model of its current location, the environment, and its goal. The robot will reactively detect and avoid obstacles and incorporate newly detected obstacles into the robot's path planning algorithms.

This framing works well for obstacles, such as a box on the ground that wasn't present the last time the robot traversed a particular hallway, but humans are a different kind of obstacle. People in an environment are there for some purpose, and their behavior is determined by their underlying intention. They may want to interact with the robot and the robot may have a goal to interact with them. Thus a social robot needs to treat navigation as a social planning problem, and in order to do so, a robot needs to reason about its own intentions with respect to the intentions of humans in the environment. This coordination of intentions is what makes the problem inherently different from reactive obstacle avoidance. In this section we detail recent computational HRI work aimed at endowing robots with the ability to reason about and produce social navigation behaviors.

In the early 2000s, a series of robotics challenges posed by the American Association for Artificial Intelligence (AAAI) elicited several large-scale efforts for robots to navigate human environments [e.g., Maxwell, 2007, Michaud et al., 2007]. These attempted to take into consideration the scenarios of meeting humans and navigating around them, but still largely modeled humans as obstacles or goals and had only rudimentary concern for social interaction.

Some recent works also deal with the issue of person tracking by treating humans as obstacles or navigation goals without much consideration for social interaction (e.g., Frintrop et al. [2010], Foka and Trahanias [2010], Jung and Sukhatme [2010]). In some cases, the focus is on detecting groups from dynamic sensor data [Lau et al., 2010]. This is extended by Bellotto and Hu [2010], combining human-tracking with person-recognition using a mixture of sensors, specifically laser range finders and cameras. Their system parses clothes as well as faces for simultaneous detection and recognition during navigation.

We refer the reader to Rios-Martinez et al. [2015], who present a thorough survey of behavioral theories and concepts that inform robot social navigation with an emphasis of proxemics. They classify social navigation research into several categories: general proxemics and approach research; unfocused interaction, such as corridor passing and avoidance; and focused interaction, including approach for interaction, dialog, people-following, and side-by-side walking.

In this section, we adopt some of the same taxonomy. We discuss representations and models for social navigation, the challenge of approaching humans, navigating side-by-side with people and following them. We conclude with work that looks at the relationship between social navigation and verbal instructions.

8.1 Representations for Human-Like and Human-Aware Navigation

A framework inspiring much of the social robot navigation literature is the Social Force Model (SFM) originally developed for analyzing human pedestrian motion [Helbing and Molnar, 1995]. The model frames pedestrians as affected by a combination of pseudo-physical forces including a controlled force in the desired movement direction, a repulsive force induced by other pedestrians, borders, and obstacles, and additional habituation-declining attractors towards incidental points of interest. Ferrer et al. [2013] present a good introduction of the SFM and a description of a method for learning the parameters for a robotic navigation system based on the SFM.

Several works look to extend SFM in a variety of ways. Shiomi et al. [2014] aim to replace regular collision avoidance among human pedestrians by achieving human-like collision avoidance based on a model of how humans behave in crowds. They track humans avoiding a robot that would not change its course to estimate the parameters of a Collision Prediction Social Force Model (CP-SFM). The goal is to make the obstacle avoidance more natural-seeming and predictable. Ratsamee et al. [2013] also present a modified SFM, adding reasoning about the human's expectation of interaction with the robot via their facial orientation. By classifying the human's intention into one of three categories (approaching the robot, avoiding the robot, or expecting the robot to avoid the human), the method generates attractive and repulsive forces for the robot's navigation system.

Other works introduce alternative representations not related to SFM. Papadakis et al. [2014] describe a system that heuristically switches between a variety of "social safe zone" models around a human (round, egg-shaped, elliptical, laterally biased) taking into account human behavior and the robot's sensor certainty. Diego and Arras [2011] suggest learning a temporal affordance map indicating a probability distribution of where people are expected to be at various points in time. They then use these in a traveling-salesman-like setting with time-dependent costs to navigate a robot to minimally interfere with the human schedule. This could be useful, for example, for a noisy cleaning robot that would disturb people while they are working in an office. Bellotto et al. [2013] propose a Qualitative Trajectory Calculus (QTC), which captures the qualitative relationships between two point agents moving on a 2D plane. The calculus represents the sign of movement of each agent with respect to the other and a categorical relative movement speed. They present several interactions represented in the calculus, and a PROLOG-like solver for QTC, implemented on a mobile robot.

Inspired by people sharing a space, Knepper and Rus [2012] provide a heuristic-based obstacle avoidance method which works in a distributed fashion for multiple robots with and without humans in the space. Based on the "other's" behavior, the robot chooses to either react alone or rely on a cooperative avoidance strategy.

8.2 Approaching Humans

A robot wanting to interact has to first detect the right human to interact with and then safely and appropriately approach the human while communicating its interaction intention.

Hanajima et al. [2005] measure people's skin conductivity (as an indicator of stress) around a robotic arm moving according to varying patterns and determine a simple movement rule with the aim of reducing stress: when a robot is in the vicinity of a human, it should move more slowly. They apply this rule on an autonomous mobile robot, slowing down when approaching a human. Meisner et al. [2008] use a similar approach to develop a Galvanic Skin Response (GSR) based algorithm to find human-friendly trajectories. Duncan and Murphy [2013] perform this kind of analysis with UAVs, investigating comfortable approach heights using biometrics and surveys, but did not find any difference between a high and low approach.

Based on surveys showing a preference for lateral, specifically right-handed approach by mobile robots [see: Dautenhahn et al., 2006], Sisbot et al. [2007] suggest the notion of a Human Aware Motion Planner (HAMP), taking into consideration and reasoning about the human context. They define a "safety grid" and a "visibility" cost grid dependent on the human's orientation and posture (sitting vs. standing). The planner also takes into consideration line of sight due to obstacles. A cost-minimizing planner then finds the most human-appropriate navigation path.

Chi-Pang et al. [2011] also develop a human-aware motion planner, expanding on Sisbot et al. [2007] by adding "harmonious rules" (e.g., no-collusion, no-interference, waiting, etc.) as well as the sensitive fields of humans and other robots, taking into consideration not only the human's position and orientation, but also the human's movement vector. Additionally, the motion planner considers social hierarchies and priorities of passing through narrow spaces.

The types of social parameters that a motion planner like HAMP and its extensions require can be learned from data. Avrunin and Simmons [2014] use human data to model socially appropriate approach paths toward an unsuspecting human based on the human's orientation with respect to the robot. Luber et al. [2012] learn prototypes of relative motions between two people by clustering sequences of top-down camera data using Dynamic Time Warping (DTW) matching. These prototypes are then classified into social contexts, modeled as appropriate angles of approach.

Others have used explicit demonstrations of proper behavior in order to estimate model parameters. Torta et al. [2011] use a teleoperated robot, allowing participants to stop the robot at "maximal", "minimal", and "optimal" distances to determine parameters for the obstacle force fields in an attractor-based navigation system. Conversely, Lichtenthäler et al. [2013] ask participants to operate a mobile robot crossing a human's path and determine that human-robot distance was the best predictor to determine the stopping condition for the robot.

A particular type of approaching behavior the research community has focused on is how a robot should pass by a person when in a narrow space, such as a hallway. Early work on this problem was framed primarily as an obstacle avoidance problem. Pacchierotti et al. [2006] refer to social distances, but mostly find that people prefer the robot to take as wide of a detour around them as possible.

Kirby et al. [2009] model the social navigation problem as a set of constraints to be weighted and optimized simultaneously. These constraints include: travel distance minimization obstacle avoidance, person avoidance (including some social aspects, such as pass-on-right and personal space), constant velocity, and inertia. Pandey and Alami [2010a] show a method to generate a smooth path by taking into account a number of sensor readings and rules, which are dynamically evaluated from sensor readings. Their system combines reasoning about global structure, local clearance, and social situations, including single and multiple people.

Approaching people in open public spaces poses additional unique questions related to the environment and the multitude of people in the scene. Kanda et al. [2009] track human behavior in a public space and classify short trajectories into styles and speeds. They then chain those to identify global behaviors. A robot uses these behaviors in an anticipatory planner to choose roaming areas and make approach decisions. Satake et al. [2013] add to this a method of specific approach trajectory including monitoring the reaction of the human approach target. Satake et al. [2014] employ a heuristic-based approach in combination with a Support Vector Machine (SVM) learner based on observer classification. They use this to model shop territory extents with the goal of respecting shopkeepers' commercial space. Similarly, Kitade et al. [2013] develop a model for shopping mall robots to find appropriate positions to wait for human shoppers.

8.3 Navigating Alongside People

Socially appropriate person-accompanying and person-following is another challenge for social navigation. In early work, Gockley et al. [2007] compare various heuristics for path-following and direction-following and found direction-following to be more natural and matching expectations. Related, Brookshire [2010] uses Histogram of Oriented Gradient (HOG) features combined with offline SVM learning of pedestrian features to inform a following algorithm that estimates the human leader's position and heading and follows at a predefined distance.

Some work tackles the harder problem of a robot following a human side-by-side [Sviestins et al., 2007]. This is more difficult than person-following for several reasons: First, sensing the human when standing side-by-side is more difficult on many platforms where sensors tend to be mounted on the robot's front. Second, navigating side-by-side requires the anticipation of future actions in order for the robot to be traveling at appropriate speeds and ready for upcoming direction changes. Researchers addressing side-by-side navigation have typically represented this as a collaborative planning problem. Morales et al. [2012] develop a collaborative model based on humans walking side-by-side. Building on that work, Murakami et al. [2014] specifically examine the situation where the robot did not know the destination of the human. They implement a model and a motion controller that switches between a collaborative state to a leader-follower state in which the robot slightly falls behind until it re-estimates the collaborative navigation goal. Kuderer and Burgard [2014] describe a leader-follower framework by also modeling human social navigation that takes the collaborative nature of the navigating pair into account. This allows the robot to compute plans that minimize the long-term deviation from the shared trajectory.

Koo and Kwon [2009] take a similar view of navigation as collaboration and have identified the need to detect the human's interaction intention when crossing paths with a robot. They use K-means clustering for human detection and a combination of Kalman filtering and Hidden Markov Models to infer the human intention. Tranberg et al. [2009] use Case-Base Reasoning to achieve a similar classification, but utilize it not for people passing, but to position the robot at the most appropriate position relative to the human. Finally, Topp and Christensen [2010] present a framework that serves for a robot to segment a space into semantic regions, which have physical features, but also make sense to humans sharing the space with the robot. They implemented a method to represent and detect regions with human labeling.

Treating navigation as a collaboration illuminates the importance of people being able to understand and interpret the robot's intentions as well (see also: Section 6.4). Several researchers have worked toward generating intent-expressive navigation for mobile robots. Kruse et al. [2014] address the issue of legibility of a dynamically re-planning robot trying to avoid moving humans in its environment. Their approach minimizes illegible erratic robot motion by allowing the robot to adjust it velocity to avoid future collisions rather than changing trajectory. Fischer et al. [2014] evaluate the role of beeping in a robot passing a human. Participants were more comfortable with a beeping robot, in particular with a rising contour beep. This has clear design implication for mobile robots moving around people. Trautman et al. [2015] present a framework for a mobile robot navigating dense human crowds, in an attempt to avoid the "Freezing Robot Problem" arising from a planner determining that all paths are unsafe. They do so by jointly modeling the human and the robot decision-making as a collaborative process, under the assumption that the human's navigation plan will also adapt to the robot's movement in the shared space.

8.4 Navigation and Verbal Instructions

Research on social navigation surveyed above focuses on inferring what a human might want the robot to do based on proxemics or other motion or attention cues given off by the human. But several works deal with navigating based on explicit human verbal instructions.

Duvallet et al. [2013] construct a policy for a mobile robot navigating unknown environments using human natural language directions. Kollar et al. [2013] use a knowledge base about the environment to ground human instructions, and also update the knowledge base from the human-robot dialog. Hemachandra et al. [2011] combine person-following with natural language processing in the context of a guided tour. The robot follows the human guide and uses the human's utterances to construct a semantic map of the toured space. Fasola and Mataric [2013] model dynamic spatial relationships representing "to", "through", and "around" and describe path generation methods based on these representations. Yuan et al. [2009] use a Markov Random Field approach in combination with a multi-level model of instruction granularity (e.g., room names vs. object locations) to identify locations from natural human instruction. This can then be used as input for a waypoint route planner.

Another way social interaction plays into robot navigation is by asking for and following human directions. The ACE system [Bauer et al., 2009a,b, Muhlbauer et al., 2009] consists of a spoken dialog system integrated with a deictic gesture detector. The robot updates its navigation planning based on verbal feedback from humans along the path. They also present a method for translating the human direction-giver's frame of reference into a global frame.

In summary, based on the rich tradition of robot navigation research, the area of social navigation is also one of the most active fields of computational HRI. In our survey we found an abundance of computational models, algorithms, and systems. Many of the works were geared toward real-world scenarios, including how to approach people in general and in public settings, and how to follow or navigate alongside people. In addition, we surveyed research combining social navigation with verbal instruction and dialog. This enables robots to learn more detailed context about the space they are navigating, navigate based on the humans' explicit intentions, and engage humans in a dialog about the navigation they are attempting.

Most of the work presented here deals with ground-based point robots, with only a few recent papers tackling social navigation of aerial vehicles (AV). This tracks the general trajectory of robotics research in the past decade, and we can expect more work in the area of social AV navigation in coming years. There has also not been much research combining navigation and subsequent verbal interaction aiming at a more complete HRI scenario, which includes meshed or sequenced approach and dialog. A navigating robot can also make use of nonverbal behaviors, both for inference and in a generative manner. We did not find much work on the combination of navigation and kinesics, vocalics, or haptics. As the field is growing, combining navigation with the computational challenges presented in other sections of this survey thereby presents a variety of promising research directions.

9

Robots Learning from Human Teachers

Machine learning (ML) is a highly active area of research in the fields of computer science and robotics. Whereas much of robotics-related ML is concerned with learning from interactions with the environment and from data sets collected offline, researchers in HRI must address the particular challenge of robots learning in real-time from human input.

The motivation for developing robots capable of learning from human teachers or demonstrators is twofold. First, there is an engineering motivation: Complex behaviors can be more readily modeled from a demonstrated behavior than formulated analytically. The second motivation has to do with realistic usability: It is impossible to pre-program robots with every needed skill or even to predict all use cases. As a result, robots need the ability to learn new skills after they are deployed.

A number of surveys have been published in the past years on the topic of robots learning from humans [Billard et al., 2008, Brenna et al., 2009, Chernova and Thomaz, 2014]. Due to the recency of these surveys, this section is less detailed than the previous sections, and the reader is encouraged to refer to the above references. In this section, we provide more of a bibliographic overview of the trends we discovered in our survey, with reference to select examples. We specifically emphasize works which frame the robot learning process as a social interaction, a framework also called Socially Guided Machine Learning [Thomaz, 2006]. Socially-guided ML is founded on the view that robots interacting with people to learn new skills should utilize social behaviors and conventions. They should participate in the teaching and learning partnership as a two-way collaboration. Moreover, it posits that the ability to leverage social skills is more than a good interface for people. It can positively impact the kinds of input the human gives as well as the underlying learning mechanisms, supporting the system's success in a real-time interactive learning session.

9.1 Characterizing the Human Learning Input

Treating the robot learning problem as a social interaction stands in contrast to the traditional ML setting. Often ML algorithms make conservative assumptions about the distribution of input data, assuming that examples are independent and identically distributed. To challenge these assumptions, several works have characterized the data provided by an end-user in the context of a social interaction with the robot.

Thomaz et al. [2006b,a] present observations about human teachers in a reinforcement learning (RL) paradigm. They find a tendency to provide guidance of future actions and less feedback for past actions, observed a positive bias in RL rewards, and also found that the human-generated reward signal changes as the learning process progresses. The positive bias was also observed by Thomaz and Cakmak [2009], showing that examples provided by humans contain more positive outcomes than those collected by a robot through random exploration or by systematically scanning the state space. Adapting their algorithms to account for these tendencies resulted in better learning. In a learning from demonstration context, Nagai et al. [2008] showed that human demonstrations towards infants tend to involve *motionese* (increased motion, exaggerations and distinct pauses) that makes it easier to segment skills and to track task-related objects.

9.2 Extending Imitation Learning

The robotics literature has traditionally framed the problem of learning from humans as *imitation learning*: The human instructor moves a robot through a trajectory, or demonstrates a set of actions that the robot then imitates. This notion is inspired by imitation capabilities in animals [Tomasello, 2001] and humans [Meltzoff, 1996]. Many examples of imitation learning in robots can be found in the surveys mentioned above [Brenna et al., 2009, Chernova and Thomaz, 2014]. Recent work on social robot learning in the HRI community often rethinks straightforward demonstrations and finds new ways to transfer tasks or skills to robots.

Cakmak et al. [2009, 2010] suggest a number of simpler social learning mechanisms that are not full-fledged imitations. These include stimulus enhancement (increasing saliency of objects used by others), mimicking (copying actions of others), and emulation (recreating the effects created by others using one's own actions). Their work shows similar learning results without the complexity of full imitation learning.

Akgun et al. [2012] proposed representing manipulation tasks with sparse keyframes instead of trajectories sampled at a high frequency. This resulted in a different way of demonstrating tasks: Rather than annotating the start and the end of a recorded arm movement, the human teacher indicates a sequence of waypoints or keyframes. In Mohammad and Nishida [2012], bottom-up saliency is used to segment unstructured demonstrations into a sequence of events in an unsupervised fashion, rather than asking the human to determine an appropriate segmentation.

Some researchers look at how to augment demonstrations with natural language instructions. In Duvallet et al. [2013] the robot learns to follow natural language directions that refer to landmarks along the way, from demonstrations of people following directions. Rybski et al. [2007] combine dialog and demonstration for teaching tasks to a mobile robot. In Mason and Lopes [2011] the user teaches a tidying-up task through natural language instructions. Based on this interaction, the robot learns the human's preference of what "tidy" looks like for them.

9.3 Social Scaffolding for Exploration

Learning from demonstrations, even when augmented as discussed in the previous section, still falls short of the collaborative nature of social learning in humans. For example, the field of “situated learning” looks at the social world of a child and how it contributes to their development. One key concept is “scaffolding”, where an adult provides support such that a child can achieve something they would not be able to accomplish independently [Vygotsky and Cole, 1978, Greenfield, 1984]. For an ML system interacting with a human who is motivated to help, social elements can greatly contribute to the success of the learning process, constraining and assisting the machine.

Along these lines, Knox et al. [2013] present a framework for interactive shaping whereby a human teaches a robot by positive or negative feedback signals, i.e., approval and disapproval of observed robot actions. Suay and Chernova [2011] combine learning from rewards with anticipatory guidance.

Inspired by work in human-human tutelage, Thomaz and Breazeal propose a framework for learning through social interaction [Lockerd and Breazeal, 2004]. The human teacher provides instructions to suggest actions to try and directs the robot’s attention towards objects by pointing at them. In addition, the robot continually communicates its internal state so as to maintain mutual beliefs with the human allowing them to appropriately guide the learning process. In Breazeal and Thomaz [2008], a robot’s exploration-exploitation behavior in a reinforcement learning framework is driven by a motivational system with a *novelty* and *mastery* drive. The human teacher scaffolds the robot’s learning by providing social context.

In the work of Argall et al. [2007], the human teacher critiques the performance of a robot policy learned from previous demonstrations. A critique is a binary label of “good” or “bad” provided by the teacher, at each time step of the robot’s execution of a movement policy. In a similar vein, Meriçli et al. [2012] involves corrective demonstrations in which the teacher proposes an alternative action to be executed in the same state or as a modification to the action previously selected by the robot.

9.4 Making the Learning Process Transparent

To support a situated learning interaction, a good instructor maintains a mental model of the learner's state—what is understood so far and what remains confusing or unknown. This helps the teacher appropriately structure the learning task with timely feedback and guidance. The learner helps the instructor by expressing their internal state via communicative acts that reveal understanding, confusion, and attention. Through this reciprocal and coupled interaction, the learner and instructor cooperate to help the instructor maintain a good mental model of the learner, and help the learner leverage from instruction to build correct models, representations, and associations. Therefore it is important for a socially interactive robot learner to communicate its internal state for the human teacher through verbal and nonverbal behaviors.

In Alexandrova et al. [2014] the robot uses visualizations to improve the human's mental model of what the robot learns from their demonstrations. This enables the teacher to correct the learned model by directly interacting with the visualization. Similarly, De Tommaso et al. [2012] devise a robot learner that has the ability to project visualizations onto the environment to aid the learning from demonstration process. Mühlig et al. [2012] use a more anthropomorphic attention mechanism to give feedback to the human teacher by gazing towards objects that are relevant for the task being demonstrated. The human can then increase an object's saliency by shaking or pointing to it and decrease the saliency by hiding it.

Active Learning is a variation on supervised learning in which the machine learner can decide which examples need to be labeled. Selecting an example to be labeled communicates to a human teacher that this instance of the problem is unknown by the current model and that labeling it would be most informative in the learning process. Thus the query itself gives information about the learned model, providing transparency into the state of the learning process. Suay et al. [2012] compared different policy learning methods and showed that people's perceived accuracy of the learner was highest when the learner was active, i.e., when the learner requested demonstrations in particular states where it was least confident. Along this line, Chernova and Veloso [2010] present an active learning approach in which the robot identifies informative states for which the human is requested to provide a demonstration. This allows a human to teach multiple robots simultaneously. Chao et al. [2010] use active learning where a robot requests labels for particular object instances. Lastly, Cakmak and Thomaz [2012] propose three types of robot questions that can be asked during demonstrations, based on observations of how humans ask questions.

In summary, we surveyed the subset of computational learning research in HRI in which the human-robot learning dyad is framed as a social interaction. We found work along several research trajectories. Some research attempts to characterize the nature of human-generated input for machine learning algorithms and subsequently develop systems making use of the human-specific features. Other socially-guided ML research works to extend traditional "direct demonstration" imitation learning by adding saliency and waypoint information, such as by way of combining verbal instruction with demonstration. Finally, there is a body of research that views interactive robot learning as a collaborative tutorial. According to that view, the human provides social scaffolding to support the learning process, and the robotic system incorporates mechanisms of communicating the learning process for instructional transparency.

Most systems discussed here were implemented and evaluated only in a laboratory setting. As robots get deployed in real-world situations, we expect new socially-guided ML systems to take advantage of the much larger quantity of interaction data from lay users in their daily environments. This could result in socially-guided ML systems that make use of learning algorithms that rely on large quantities of training data, such as convolutional neural networks, which have provided impressive results in other fields of ML in recent years. The combination of real-world deployment and large databases will certainly shed new light on how robots can learn in a real-world social environment, a largely unexplored challenge in the socially-guided ML literature, and one that offers a promising next step for the community.

10

Conclusion

This paper reviewed approximately ten years of computational research in HRI, covering a period in which the field has experienced significant growth. This expansion was evident in bibliometric trends uncovered in the twelve venues we surveyed. In particular, we saw a steady increase in HRI research in broad, general-interest, and established robotics venues.

As computational HRI research has been more accepted into the wider robotics community, specialized HRI venues have reacted by publishing a larger percentage of empirical and social-science oriented work, struggling to maintain a balance between the two. The community has attempted to reclaim this balance by including specialized computational and systems-oriented tracks and submission areas in HRI conferences as well as calling for computational and systems-oriented special issues in its journals.

Our literature review shows that computational HRI research builds on two foundations: traditional robotics research and research on human social behavior and interaction. This is reflected in the categories we identified and in the structure of the paper. Each section builds on established robotics fields—perception, navigation, learning, planning, and manipulation—but views them through the prism of human social capacities, activities, and behaviors.

Using predetermined inclusion criteria, we found the available work to be broad, but often biased toward popular subtopics. Sometimes these correspond to highly pressing problems in human-robot social interactions, in other cases they reveal low-hanging fruit, inspired by available sensor technology and established algorithms in robotics. For example, the work on human perception has a large component of attention and engagement detection, which is a core challenge in HRI, but it also deals widely with activity inference, a traditionally active area in the general computer science literature.

We divided the research into two segments: foundations and high-level competencies. Foundations include perceiving human activity, the verbal and nonverbal interaction modalities, and affective computing aspects of HRI. High-level competencies build on these and include intention recognition, collaboration, navigation, and learning.

The work on verbal communication covered both content and paralinguistics, with an emphasis on the embodied, real-time, and situated nature of robotics. Algorithms for nonverbal behavior covered a range of body movements (kinesics) and spatial movement (proxemics), but we found relatively little new work on touch interaction. Affective computing research for HRI was found to be a highly active subfield, and—in the eyes of the general public—often synonymous with social robotics and HRI. We surveyed work on models of emotions, expression, and detection of emotional behavior.

Building on the foundation of intention recognition, we continued to survey the research challenges of social collaboration, navigation, and learning. These can be seen as HRI extensions of traditional robotics fields. All three are vast research areas of computational HRI. We tried to focus on the work that tackles challenges unique to the social context and had to exclude many papers which dealt with human interaction, but did not include a prominent social component. For collaboration, this meant focusing on the work dealing with cognitive and executive frameworks, timing, and the social aspects of handovers and shared manipulation. In navigation, we focused on work that treated humans as more than mere obstacles, and that dealt with social spatial interactions such as side-by-side navigation and approaching humans in a socially appropriate manner. The learning literature surveyed took a particular look at learning as a social activity rather than at learning from human demonstration.

Any research survey of this scope is inherently limited, excludes excellent work, and cannot be considered complete. In addition, our choice of categorization leads to inclusion and exclusion of research that would have been considered differently under an alternative taxonomy. We nonetheless believe that the above survey paints a relatively representative picture of the current state of the art in computational HRI research, and as such can be a valuable reference for moving forward in this field.

References

H. Admoni, B. Hayes, D. Feil-Seifer, D. Ullman, and B. Scassellati. Are you looking at me?: perception of robot attention is mediated by gaze type and group size. In *Proceedings of the 8th ACM/IEEE international conference on human-robot interaction*, pages 389–396. IEEE Press, 2013.

H. Admoni, A. Dragan, S. S. Srinivasa, and B. Scassellati. Deliberate delays during robot-to-human handovers improve compliance with gaze communication. In *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*, pages 49–56. ACM, 2014.

H. S. Ahn and J. Y. Choi. Emotional behavior decision model based on linear dynamic systems for intelligent service robots. In *Robot and Human interactive Communication, 2007. RO-MAN 2007. The 16th IEEE International Symposium on*, pages 786–791. IEEE, 2007.

H. S. Ahn, D.-W. Lee, D. Choi, D.-Y. Lee, M. Hur, and H. Lee. Appropriate emotions for facial expressions of 33-dofs android head ever-4 h33. In *RO-MAN, 2012 IEEE*, pages 1115–1120. IEEE, 2012.

B. Akgun, M. Cakmak, J. W. Yoo, and A. L. Thomaz. Trajectories and keyframes for kinesthetic teaching: A human-robot interaction perspective. In *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*, pages 391–398. ACM, 2012.

S. Al Moubayed, M. Baklouti, M. Chetouani, T. Dutoit, A. Mahdhaoui, J-C. Martin, S. Ondas, C. Pelachaud, J. Urbain, and M. Yilmaz. Generating robot/agent backchannels during a storytelling experiment. In *Robotics and Automation, 2009. ICRA'09. IEEE International Conference on*, pages 3749–3754. IEEE, 2009.

R. Alazrai and C. S. G. Lee. Real-time emotion identification for socially intelligent robots. In *Robotics and Automation (ICRA), 2012 IEEE International Conference on*, pages 4106–4111. IEEE, 2012.

J. Aleotti, V. Micelli, and S. Caselli. An affordance sensitive system for robot to human object handover. *International Journal of Social Robotics*, 6(4): 653–666, 2014.

S. Alexandrova, M. Cakmak, K. Hsiao, and L. Takayama. Robot programming by demonstration with interactive action visualizations. *Proceedings of Robotics: Science and Systems, Berkeley, USA*, 2014.

A. Aly and A. Tapus. A model for synthesizing a combined verbal and nonverbal behavior based on personality traits in human-robot interaction. In *Proceedings of the 8th ACM/IEEE international conference on Human-robot interaction*, pages 325–332. IEEE Press, 2013.

M. Ammi, V. Demulier, S. Caillou, Y. Gaffary, Y. Tsalamlal, J.-C. Martin, and A. Tapus. Haptic human-robot affective interaction in a handshaking social protocol. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, pages 263–270. ACM, 2015.

S. Andrist, E. Spannan, and B. Mutlu. Rhetorical robots: making robots more effective speakers using linguistic cues of expertise. In *Proceedings of the 8th ACM/IEEE international conference on Human-robot interaction*, pages 341–348. IEEE Press, 2013.

S. Andrist, X. Z. Tan, M. Gleicher, and B. Mutlu. Conversational gaze aversion for humanlike robots. In *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*, pages 25–32. ACM, 2014.

M. L. Anjum, O. Ahmad, S. Rosa, J. Yin, and B. Bona. Skeleton tracking based complex human activity recognition using kinect camera. In *Social Robotics*, pages 23–33. Springer, 2014.

B. Argall, B. Browning, and M. Veloso. Learning by demonstration with critique from a human teacher. In *Proceedings of the ACM/IEEE international conference on Human-robot interaction*, pages 57–64. ACM, 2007.

M. Argyle and J. Dean. Eye-contact, distance and affiliation. *Sociometry*, pages 289–304, 1965.

M. Argyle, R. Ingham, F. Alkema, and M. McCallin. The different functions of gaze. *Semiotica*, 7(1):19–32, 1973.

L. Aryananda. Learning to recognize familiar faces in the real world. In *Robotics and Automation, 2009. ICRA'09. IEEE International Conference on*, pages 1991–1996. IEEE, 2009.

E. Avrunin and R. Simmons. Socially-appropriate approach paths using human data. In *Robot and Human Interactive Communication, 2014 RO-MAN: The 23rd IEEE International Symposium on*, pages 1037–1042, Aug 2014.

E. Avrunin, J. Hart, A. Douglas, and B. Scassellati. Effects related to synchrony and repertoire in perceptions of robot dance. In *Proceedings of the 6th international conference on Human-robot interaction*, pages 93–100. ACM, 2011.

D. A. Baldwin, J. A. Baird, M. M. Saylor, and M. A. Clark. Infants parse dynamic action. *Child Development*, 72(3):708–717, 2001. ISSN 1467-8624. URL `http://dx.doi.org/10.1111/1467-8624.00310`.

L. Bascetta, G. Ferretti, P. Rocco, H. Ardö, H. Bruyninckx, E. Demeester, and E. D. Lello. Towards safe human-robot interaction in robotic cells: an approach based on visual tracking and intention estimation. In *Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on*, pages 2971–2978. IEEE, 2011.

A. Bauer, K. Klasing, G. Lidoris, Q. Mühlbauer, F. Rohrmüller, S. Sosnowski, T. Xu, K. Kühnlenz, D. Wollherr, and M. Buss. The autonomous city explorer: Towards natural human-robot interaction in urban environments. *International Journal of Social Robotics*, 1(2):127–140, 2009a.

A. Bauer, D. Wollherr, and M. Buss. Information retrieval system for human-robot communication - asking for directions. In *Robotics and Automation, 2009. ICRA '09. IEEE International Conference on*, pages 4150–4155, May 2009b.

M. Beetz, F. Stulp, P. Esden-Tempski, A. Fedrizzi, U. Klank, I. Kresse, A. Maldonado, and F. Ruiz. Generality and legibility in mobile manipulation. *Autonomous Robots*, 28(1):21–44, 2010.

N. Bellotto and H. Hu. A bank of unscented kalman filters for multimodal human perception with mobile service robots. *International Journal of Social Robotics*, 2(2):121–136, 2010.

N. Bellotto, M. Hanheide, and N. Van de Weghe. Qualitative design and implementation of human-robot spatial interactions. In *Social Robotics*, pages 331–340. Springer, 2013.

H. Ben Amor, G. Neumann, S. Kamthe, O. Kroemer, and J. Peters. Interaction primitives for human-robot cooperation tasks. In *Robotics and Automation (ICRA), 2014 IEEE International Conference on*, pages 2831–2837. IEEE, 2014.

C. C. Bennett and S. Šabanović. Deriving minimal features for human-like facial expressions in robotic faces. *International Journal of Social Robotics*, 6(3):367–381, 2014.

M. Berlin, J. Gray, A. L. Thomaz, and C. Breazeal. Perspective taking: An organizing principle for learning in human-robot interaction. In *Proceedings of the National Conference on Artificial Intelligence*, volume 21(2), page 1444. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, 2006.

C. Bevan and D. Stanton Fraser. Shaking hands and cooperation in telepresent human-robot negotiation. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, pages 247–254. ACM, 2015.

A. Billard, S. Callinon, R. Dillmann, and S. Schaal. Robot programming by demonstration. In B. Siciliano and O. Khatib, editors, *Handbook of Robotics*, chapter 59. Springer, New York, NY, USA, 2008.

S. N. Blisard and M. Skubic. Modeling spatial referencing language for human-robot interaction. In *Robot and Human Interactive Communication, 2005. ROMAN 2005. IEEE International Workshop on*, pages 698–703. IEEE, 2005.

S. Boucenna, P. Gaussier, P. Andry, and L. Hafemeister. A robot learns the facial expressions recognition and face/non-face discrimination through an imitation game. *International Journal of Social Robotics*, 6(4):633–652, 2014.

M. E. Bratman. Shared cooperative activity. *The philosophical review*, 101 (2):327–341, 1992.

C. Breazeal and B. Scassellati. How to build robots that make friends and influence people. In *Intelligent Robots and Systems, 1999. IROS'99. Proceedings. 1999 IEEE/RSJ International Conference on*, volume 2, pages 858–863. IEEE, 1999.

C. Breazeal and A. L. Thomaz. Learning from human teachers with socially guided exploration. In *Robotics and Automation, 2008. ICRA 2008. IEEE International Conference on*, pages 3539–3544. IEEE, 2008.

C. Breazeal, A. Wang, and R. Picard. Experiments with a robotic computer: body, affect and cognition interactions. In *Human-Robot Interaction (HRI), 2007 2nd ACM/IEEE International Conference on*, pages 153–160. IEEE, 2007.

C. Breazeal, A. Takanishi, and T. Kobayashi. Social robots that interact with people. In *Springer handbook of robotics*, pages 1349–1369. Springer, 2008.

C. Breazeal, J. Gray, and M. Berlin. An embodied cognition approach to mindreading skills for socially intelligent robots. *The International Journal of Robotics Research*, 28(5):656–680, 2009.

C. L. Breazeal. *Designing sociable robots*. MIT press, 2004.

P. Bremner and U. Leonards. Speech and gesture emphasis effects for robotic and human communicators: A direct comparison. In *HRI*, pages 255–262, 2015.

D. A. Brenna, C. Sonia, V. Manuela, and B. Brett. A survey of robot learning from demonstration. *Robotics and Autonomous Systems*, 57(5), 2009.

M. Bretan, G. Hoffman, and G. Weinberg. Emotionally expressive dynamic physical behaviors in robots. *International Journal of Human-Computer Studies*, 78:1–16, 2015.

A. G. Brooks and R. C. Arkin. Behavioral overlays for non-verbal communication expression on a humanoid robot. *Autonomous Robots*, 22(1):55–74, 2007.

A. G. Brooks and C. Breazeal. Working with robots and objects: Revisiting deictic reference for achieving spatial common ground. In *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*, pages 297–304. ACM, 2006.

J. Brookshire. Person following using histograms of oriented gradients. *International journal of social robotics*, 2(2):137–146, 2010.

X. Broquère, A. Finzi, J. Mainprice, S. Rossi, D. Sidobre, and M. Staffa. An attentional approach to human–robot interactive manipulation. *International Journal of Social Robotics*, 6(4):533–553, 2014.

B. Burger, I. Ferrané, F. Lerasle, and G. Infantes. Two-handed gesture recognition and fusion with speech to command a robot. *Autonomous Robots*, 32(2):129–147, 2012.

J. Butterfield, O. C. Jenkins, D. M. Sobel, and J. Schwertfeger. Modeling aspects of theory of mind with markov random fields. *International Journal of Social Robotics*, 1(1):41–51, 2009.

M. Cakmak and A. L. Thomaz. Designing robot learners that ask good questions. In *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*, pages 17–24. ACM, 2012.

M. Cakmak, N. DePalma, A. L. Thomaz, and R. Arriaga. Effects of social exploration mechanisms on robot learning. In *Robot and Human Interactive Communication, 2009. RO-MAN 2009. The 18th IEEE International Symposium on*, pages 128–134. IEEE, 2009.

M. Cakmak, N. DePalma, R. I. Arriaga, and A. L. Thomaz. Exploiting social partners in robot learning. *Autonomous Robots*, 29(3-4):309–329, 2010.

M. Cakmak, S. S. Srinivasa, M. K. Lee, S. Kiesler, and J. Forlizzi. Using spatial and temporal contrast for fluent robot-human hand-overs. In *Proceedings of the 6th international conference on Human-robot interaction*, pages 489–496. ACM, 2011.

R. Cantrell, M. Scheutz, P. Schermerhorn, and X. Wu. Robust spoken instruction understanding for hri. In *Human-Robot Interaction (HRI), 2010 5th ACM/IEEE International Conference on*, pages 275–282. IEEE, 2010.

R. Cantrell, P. Schermerhorn, and M. Scheutz. Learning actions from human-robot dialogues. In *RO-MAN, 2011 IEEE*, pages 125–130. IEEE, 2011.

P. Carcagnì, D. Cazzato, M. Del Coco, M. Leo, G. Pioggia, and C. Distante. Real-time gender based behavior system for human-robot interaction. In *Social Robotics*, pages 74–83. Springer, 2014.

M. Carpenter, K. Nagell, Tomasello, G. M. Butterworth, and C. Moore. Social cognition, joint attention, and communcative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development*, 63(4):1–174, 1998.

J. Cassell. *Embodied conversational agents*. MIT press, 2000.

W. P. Chan, C. A. C. Parker, H. M. Van Der Loos, and E. A. Croft. A human-inspired object handover controller. *The International Journal of Robotics Research*, 32(8):971–983, 2013.

C. Chao and A. L. Thomaz. Timing in multimodal turn-taking interactions: Control and analysis using timed petri nets. *Journal of Human-Robot Interaction*, 1(1), 2012.

C. Chao and A. L. Thomaz. Controlling social dynamics with a parametrized model of floor regulation. *Journal of Human-Robot Interaction*, 2(1):4–19, 2013.

C. Chao, M. Cakmak, and A. L. Thomaz. Transparent active learning for robots. In *Human-Robot Interaction (HRI), 2010 5th ACM/IEEE International Conference on*, pages 317–324. IEEE, 2010.

C. Chao, J. Lee, M. Begum, and A. L. Thomaz. Simon plays simon says: The timing of turn-taking in an imitation game. In *RO-MAN, 2011 IEEE*, pages 235–240. IEEE, 2011.

S. Chao, S. Masahiro, S. Christian, K. Takayuki, and I. Hiroshi. A model of distributional handing interaction for a mobile robot. In *Proceedings of Robotics: Science and Systems*, Berlin, Germany, June 2013.

T. L. Chen and C. C. Kemp. Lead me by the hand: Evaluation of a direct physical interface for nursing assistant robots. In *Proceedings of the 5th ACM/IEEE international conference on Human-robot interaction*, pages 367–374. IEEE Press, 2010.

S. Chernova and A. L. Thomaz. Robot learning from human teachers. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, 8(3):1–121, 2014. URL `http://dx.doi.org/10.2200/S00568ED1V01Y201402AIM028`.

S. Chernova and M. Veloso. Confidence-based multi-robot learning from demonstration. *International Journal of Social Robotics*, 2(2):195–215, 2010.

L. Chi-Pang, C. Chen-Tun, C. Kuo-Hung, and F. Li-Chen. Human-centered robot navigation: Towards a harmoniously human-robot coexisting environment. *Robotics, IEEE Transactions on*, 27(1):99–112, Feb 2011. ISSN 1552-3098.

V. Chidambaram, Y.-H. Chiang, and B. Mutlu. Designing persuasive robots: how robots might persuade people using vocal and nonverbal cues. In *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*, pages 293–300. ACM, 2012.

A. Chrungoo, S. S. Manimaran, and B. Ravindran. Activity recognition for natural human robot interaction. In *Social Robotics*, pages 84–94. Springer, 2014.

V. Chu, K. Bullard, and A. L. Thomaz. Multimodal real-time contingency detection for hri. In *Intelligent Robots and Systems (IROS 2014), 2014 IEEE/RSJ International Conference on*, pages 3327–3332. IEEE, 2014.

Y. Chuang, L. Chen, G. Zhao, and G. Chen. Hand posture recognition and tracking based on bag-of-words for human robot interaction. In *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, pages 538–543. IEEE, 2011.

F. Cid, J. A. Prado, P. Bustos, and P. Nunez. A real time and robust facial expression recognition and imitation approach for affective human-robot interaction using gabor filtering. In *Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on*, pages 2188–2193. IEEE, 2013.

H. H. Clark. *Using language*. Cambridge university press, 1996.

P. R. Cohen and H. J. Levesque. Teamwork. *Nous*, 25(4):487–512, 1991.

S. Costa, F. Soares, and C. Santos. Facial expressions and gestures to convey emotions with a humanoid robot. In *Social Robotics*, pages 542–551. Springer, 2013.

C. Darwin. *The expression of the emotions in man and animals.* John Murray, 1873.

K. Dautenhahn, M. Walters, S. Woods, K. L. Koay, C. L. Nehaniv, A. Sisbot, R. Alami, and T. Siméon. How may i serve you?: A robot companion approaching a seated person in a helping context. In *Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-robot Interaction*, HRI '06, pages 172–179, New York, NY, USA, 2006. ACM. ISBN 1-59593-294-1. URL http://doi.acm.org.ezprimo1.idc.ac.il/10.1145/1121241.1121272.

D. De Tommaso, S. Calinon, and D. G. Caldwell. A tangible interface for transferring skills. *International Journal of Social Robotics*, 4(4):397–408, 2012.

R. Deits, S. Tellex, P. Thaker, D. Simeonov, T. Kollar, and N. Roy. Clarifying commands with information-theoretic human-robot dialog. *Journal of Human-Robot Interaction*, 2(2):58–79, 2013.

D. C. Dennett. *The Intentional Stance.* MIT Press, 1989.

G. Diego and K. O. Arras. Please do not disturb! minimum interference coverage for social robots. In *Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on*, pages 1968–1973, Sept 2011.

H. Ding, G. Reißig, K. Wijaya, D. Bortot, K. Bengler, and O. Stursberg. Human arm motion modeling and long-term prediction for safe and efficient human-robot-interaction. In *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, pages 5875–5880. IEEE, 2011.

A. Dragan, R. Holladay, and S. Srinivasa. Deceptive robot motion: synthesis, analysis and experiments. *Autonomous Robots*, 39(3):331–345, 2015.

A. D. Dragan, K. C. T. Lee, and S. S. Srinivasa. Legibility and predictability of robot motion. In *Human-Robot Interaction (HRI), 2013 8th ACM/IEEE International Conference on*, pages 301–308. IEEE, 2013.

D. Droeschel, J. Stückler, and S. Behnke. Learning to interpret pointing gestures with a time-of-flight camera. In *Proceedings of the 6th international conference on Human-robot interaction*, pages 481–488. ACM, 2011a.

D. Droeschel, J. Stückler, D. Holz, and S. Behnke. Towards joint attention for a domestic service robot-person awareness and gesture recognition using time-of-flight cameras. In *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, pages 1205–1210. IEEE, 2011b.

B. A. Duncan and R. R. Murphy. Comfortable approach distance with small unmanned aerial vehicles. In *RO-MAN, 2013 IEEE*, pages 786–792, Aug 2013.

F. Duvallet, T. Kollar, and A. Stentz. Imitation learning for natural language direction following through unknown environments. In *Robotics and Automation (ICRA), 2013 IEEE International Conference on*, pages 1047–1053. IEEE, 2013.

A. Edsinger and C. C Kemp. Human-robot interaction for cooperative manipulation: Handing objects to one another. In *Robot and Human interactive Communication, 2007. RO-MAN 2007. The 16th IEEE International Symposium on*, pages 1167–1172. IEEE, 2007.

P. Ekman and W. V. Friesen. The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *Semiotica*, 1(1):49–98, 1969.

P. Evrard and A. Kheddar. Homotopy-based controller for physical human-robot interaction. In *Robot and Human Interactive Communication, 2009. RO-MAN 2009. The 18th IEEE International Symposium on*, pages 1–6. IEEE, 2009.

R. Fang, M. Doering, and J. Y. Chai. Embodied collaborative referring expression generation in situated human-robot interaction. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, pages 271–278. ACM, 2015.

J. Fasola and M. J. Mataric. Modeling dynamic spatial relations with global properties for natural language-based human-robot interaction. In *RO-MAN, 2013 IEEE*, pages 453–460, Aug 2013.

J. Fasola and M. J. Mataric. Interpreting instruction sequences in spatial language discourse with pragmatics towards natural human-robot interaction. In *Robotics and Automation (ICRA), 2014 IEEE International Conference on*, pages 2720–2727. IEEE, 2014.

D. Feil-Seifer and M. Mataric. Automated detection and classification of positive vs. negative robot interactions with children with autism using distance-based features. In *Proceedings of the 6th international conference on Human-robot interaction*, pages 323–330. ACM, 2011.

D. Feil-Seifer and M. J. Matarić. A multi-modal approach to selective interaction in assistive domains. In *Robot and Human Interactive Communication, 2005. ROMAN 2005. IEEE International Workshop on*, pages 416–421. IEEE, 2005.

F. Ferland, A. Aumont, D. Létourneau, and F. Michaud. Taking your robot for a walk: Force-guiding a mobile robot using compliant arms. In *Proceedings of the 8th ACM/IEEE international conference on Human-robot interaction*, pages 309–316. IEEE Press, 2013.

G. Ferrer, A. Garrell, and A. Sanfeliu. Robot companion: A social-force based approach with human awareness-navigation in crowded environments. In *Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on*, pages 1688–1694, Nov 2013.

M. Finke, K. L. Koay, K. Dautenhahn, C. L. Nehaniv, M. L. Walters, and J. Saunders. Hey, i'm over here-how can a robot attract people's attention? In *Robot and Human Interactive Communication, 2005. ROMAN 2005. IEEE International Workshop on*, pages 7–12. IEEE, 2005.

K. Fischer, L. C. Jensen, and L. Bodenhagen. To beep or not to beep is not the whole question. In *Social Robotics*, pages 156–165. Springer, 2014.

A. Flagg and K. MacLean. Affective touch gesture recognition for a furry zoomorphic machine. In *Proceedings of the 7th International Conference on Tangible, Embedded and Embodied Interaction*, pages 25–32. ACM, 2013.

A. F. Foka and P. E. Trahanias. Probabilistic autonomous robot navigation in dynamic environments with human motion prediction. *International Journal of Social Robotics*, 2(1):79–94, 2010.

T. Fong, I. Nourbakhsh, and K. Dautenhahn. A survey of socially interactive robots. *Robotics and Autonomous Systems*, 42(3-4):143–166, March 2003. ISSN 09218890. URL `http://linkinghub.elsevier.com/retrieve/pii/S092188900200372X`.

M. E. Foster, E. G. Bard, M. Guhe, R. L. Hill, J. Oberlander, and A. Knoll. The roles of haptic-ostensive referring expressions in cooperative, task-based human-robot dialogue. In *Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction*, pages 295–302. ACM, 2008.

B. Fransen, V. Morariu, E. Martinson, S. Blisard, M. Marge, S. Thomas, A. Schultz, and D. Perzanowski. Using vision, acoustics, and natural language for disambiguation. In *Proceedings of the ACM/IEEE international conference on Human-robot interaction*, pages 73–80. ACM, 2007.

S. Frintrop, A. Königs, F. Hoeller, and D. Schulz. A component-based approach to visual person tracking from a mobile platform. *International Journal of Social Robotics*, 2(1):53–62, 2010.

M. J. Gielniak and A. L. Thomaz. Generating anticipation in robot motion. In *RO-MAN, 2011 IEEE*, pages 449–454. IEEE, 2011a.

M. J. Gielniak and A. L. Thomaz. Spatiotemporal correspondence as a metric for human-like robot motion. In *Proceedings of the 6th international conference on Human-robot interaction*, pages 77–84. ACM, 2011b.

M. J. Gielniak and A. L. Thomaz. Enhancing interaction through exaggerated motion synthesis. In *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*, pages 375–382. ACM, 2012.

M. J. Gielniak, C. K. Liu, and A. L. Thomaz. Generating human-like motion for robots. *The International Journal of Robotics Research*, 32(11):1275–1301, 2013.

R. Gockley, R. Simmons, and J. Forlizzi. Modeling affect in socially interactive robots. In *Robot and Human Interactive Communication, 2006. ROMAN 2006. The 15th IEEE International Symposium on*, pages 558–563. IEEE, 2006.

R. Gockley, J. Forlizzi, and R. Simmons. Natural person-following behavior for social robots. In *Proceedings of the ACM/IEEE International Conference on Human-robot Interaction*, HRI '07, pages 17–24, New York, NY, USA, 2007. ACM. ISBN 978-1-59593-617-2. URL `http://doi.acm.org.ezprimo1.idc.ac.il/10.1145/1228716.1228720`.

R. Gomez, T. Kawahara, K. Nakamura, and K. Nakadai. Multi-party human-robot interaction with distant-talking speech recognition. In *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*, pages 439–446. ACM, 2012.

M. A. Goodrich and A. C. Schultz. Human-Robot Interaction: A Survey. *Foundations and Trends® in Human-Computer Interaction*, 1(3):203–275, February 2007. ISSN 1551-3955.

A. Gopnik, D. Sobel, L. Schulz, and C. Glymour. 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, 2001.

J. Gray and C. Breazeal. Manipulating mental states through physical action. *International Journal of Social Robotics*, 6(3):315–327, 2014.

P. M. Greenfield. I of the teacher in learning activities of everyday life. In B. Rogoff and J. Lave, editors, *Everyday cognition: its development in social context.* Harvard University Press, Cambridge, MA, 1984.

E. C. Grigore, K. Eder, A. G. Pipe, C. Melhuish, and U. Leonards. Joint action understanding improves robot-to-human object handover. In *Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on*, pages 4622–4629. IEEE, 2013.

S. Guadarrama, L. Riano, D. Golland, D. Gouhring, Y. Jia, D. Klein, P. Abbeel, and T. Darrell. Grounding spatial relations for human-robot interaction. In *Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on*, pages 1640–1647. IEEE, 2013.

G. György, N. Zoltan, C. Gergely, and B. Szilvia. Taking the intentional stance at 12 months of age. *Cognition*, 56(2):165 – 193, 1995.

J. Ham, R. H. Cuijpers, and J.-J. Cabibihan. 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, 2015.

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

S. Handri, S. Nomura, and K. Nakamura. Determination of age and gender based on features of human motion using adaboost algorithms. *International Journal of Social Robotics*, 3(3):233–241, 2011.

M. Hanheide, S. Wrede, C. Lang, and G. Sagerer. Who am i talking with? a face memory for social robots. In *Robotics and Automation, 2008. ICRA 2008. IEEE International Conference on*, pages 3660–3665. IEEE, 2008.

M. Häring, J. Eichberg, and E. André. Studies on grounding with gaze and pointing gestures in human-robot-interaction. In *Social Robotics*, pages 378–387. Springer, 2012.

T. Hashimoto, S. Hiramatsu, T. Tsuji, and H. Kobayashi. Realization and evaluation of realistic nod with receptionist robot SAYA. In *Robot and Human interactive Communication, 2007. RO-MAN 2007. The 16th IEEE International Symposium on*, pages 326–331. IEEE, 2007.

K. P. Hawkins, N. Vo, S. Bansal, and A. F. Bobick. Probabilistic human action prediction and wait-sensitive planning for responsive human-robot collaboration. In *Humanoid Robots (Humanoids), 2013 13th IEEE-RAS International Conference on*, pages 499–506. IEEE, 2013.

K. P. Hawkins, S. Bansal, N. N. Vo, and A. F. Bobick. Anticipating human actions for collaboration in the presence of task and sensor uncertainty. In *Robotics and Automation (ICRA), 2014 IEEE International Conference on*, pages 2215–2222. IEEE, 2014.

D. Helbing and P. Molnar. Social force model for pedestrian dynamics. *Physical review E*, 51(5):4282, 1995.

S. Hemachandra, T. Kollar, N. Roy, and S. Teller. Following and interpreting narrated guided tours. In *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, pages 2574–2579, May 2011.

S. Hemachandra, M. R. Walter, S. Tellex, and S. Teller. Learning spatial-semantic representations from natural language descriptions and scene classifications. In *Robotics and Automation (ICRA), 2014 IEEE International Conference on*, pages 2623–2630. IEEE, 2014.

J. Hirth, N. Schmitz, and K. Berns. Towards social robots: Designing an emotion-based architecture. *International Journal of Social Robotics*, 3(3): 273–290, 2011.

M. A. T. Ho, Y. Yamada, and Y. Umetani. An adaptive visual attentive tracker for human communicational behaviors using hmm-based td learning with new state distinction capability. *Robotics, IEEE Transactions on*, 21 (3):497–504, 2005.

J. R. Hoare and L. E. Parker. Using on-line conditional random fields to determine human intent for peer-to-peer human robot teaming. In *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*, pages 4914–4921. IEEE, 2010.

G. Hoffman. Dumb robots, smart phones: A case study of music listening companionship. In *RO-MAN, 2012 IEEE*, pages 358–363. IEEE, 2012.

G. Hoffman. Evaluating fluency in human-robot collaboration. In *Robotics: Science and Systems (RSS'13) Workshop on Human-Robot Collaboration*, 2013.

G. Hoffman and C. Breazeal. Collaboration in human-robot teams. In *Proc. of the AIAA 1st Intelligent Systems Technical Conference, Chicago, IL, USA*, 2004.

G. Hoffman and C. Breazeal. Cost-based anticipatory action selection for human–robot fluency. *Robotics, IEEE Transactions on*, 23(5):952–961, 2007.

G. Hoffman and C. Breazeal. Effects of anticipatory perceptual simulation on practiced human-robot tasks. *Autonomous Robots*, 28(4):403–423, 2010.

G. Hoffman and K. Vanunu. Effects of robotic companionship on music enjoyment and agent perception. In *Human-Robot Interaction (HRI), 2013 8th ACM/IEEE International Conference on*, pages 317–324. IEEE, 2013.

G. Hoffman and G. Weinberg. Interactive improvisation with a robotic marimba player. *Autonomous Robots*, 31(2-3):133–153, 2011.

R. M. Holladay, A. D. Dragan, and S. S. Srinivasa. Legible robot pointing. In *Robot and Human Interactive Communication, 2014 RO-MAN: The 23rd IEEE International Symposium on*, pages 217–223. IEEE, 2014.

T. M. Howard, S. Tellex, and N. Roy. A natural language planner interface for mobile manipulators. In *Robotics and Automation (ICRA), 2014 IEEE International Conference on*, pages 6652–6659. IEEE, 2014.

C.-M. Huang and A. L. Thomaz. Effects of responding to, initiating and ensuring joint attention in human-robot interaction. In *RO-MAN, 2011 IEEE*, pages 65–71. IEEE, 2011.

C.-M. Huang, M. Cakmak, and B. Mutlu. Adaptive coordination strategies for human-robot handovers. In *Proceedings of Robotics: Science and Systems*, 2015.

Chien-Ming Huang and Bilge Mutlu. Modeling and evaluating narrative gestures for humanlike robots. In *Proceedings of Robotics: Science and Systems*, Berlin, Germany, June 2013.

M. Huber, H. Radrich, C. Wendt, M. Rickert, A. Knoll, T. Brandt, and S. Glasauer. Evaluation of a novel biologically inspired trajectory generator in human-robot interaction. In *Robot and Human Interactive Communication, 2009. RO-MAN 2009. The 18th IEEE International Symposium on*, pages 639–644. IEEE, 2009.

S. Iengo, S. Rossi, M. Staffa, and A. Finzi. Continuous gesture recognition for flexible human-robot interaction. In *Robotics and Automation (ICRA), 2014 IEEE International Conference on*, pages 4863–4868. IEEE, 2014.

T. Iio, M. Shiomi, K. Shinozawa, T. Miyashita, T. Akimoto, and N. Hagita. Lexical entrainment in human-robot interaction: can robots entrain human vocabulary? In *Intelligent Robots and Systems, 2009. IROS 2009. IEEE/RSJ International Conference on*, pages 3727–3734. IEEE, 2009.

T. Iio, M. Shiomi, K. Shinozawa, T. Akimoto, K. Shimohara, and N. Hagita. Entrainment of pointing gestures by robot motion. In *Social Robotics*, pages 372–381. Springer, 2010.

C. T. Ishi, C. Liu, H. Ishiguro, and N. Hagita. Head motions during dialogue speech and nod timing control in humanoid robots. In *Proceedings of the 5th ACM/IEEE international conference on Human-robot interaction*, pages 293–300. IEEE Press, 2010.

O. C. Jenkins, G. González, and M. M. Loper. Tracking human motion and actions for interactive robots. In *Proceedings of the ACM/IEEE international conference on Human-robot interaction*, pages 365–372. ACM, 2007.

D. O. Johnson, R. H. Cuijpers, and D. van der Pol. Imitating human emotions with artificial facial expressions. *International Journal of Social Robotics*, 5(4):503–513, 2013.

B. Jung and G. S Sukhatme. Real-time motion tracking from a mobile robot. *International Journal of Social Robotics*, 2(1):63–78, 2010.

T. Kanda, D. F. Glas, M. Shiomi, and N. Hagita. Abstracting people's trajectories for social robots to proactively approach customers. *Robotics, IEEE Transactions on*, 25(6):1382–1396, Dec 2009. ISSN 1552-3098.

M. Karg, M. Schwimmbeck, K. Kuhnlenz, and M. Buss. Towards mapping emotive gait patterns from human to robot. In *RO-MAN, 2010 IEEE*, pages 258–263. IEEE, 2010.

Y. Kato, T. Kanda, and H. Ishiguro. May i help you?: design of human-like polite approaching behavior. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, pages 35–42. ACM, 2015.

J. Kedzierski, R. Muszyński, C. Zoll, A. Oleksy, and M. Frontkiewicz. Emys—emotive head of a social robot. *International Journal of Social Robotics*, 5 (2):237–249, 2013.

R. Kelley, A. Tavakkoli, C. King, M. Nicolescu, M. Nicolescu, and G. Bebis. Understanding human intentions via hidden markov models in autonomous mobile robots. In *Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction*, pages 367–374. ACM, 2008.

W. G. Kennedy, M. D. Bugajska, A. M. Harrison, and J. G. Trafton. "like-me" simulation as an effective and cognitively plausible basis for social robotics. *International Journal of Social Robotics*, 1(2):181–194, 2009.

H.-R. Kim, K. W. Lee, and D.-S. Kwon. Emotional interaction model for a service robot. In *Robot and Human Interactive Communication, 2005. ROMAN 2005. IEEE International Workshop on*, pages 672–678. IEEE, 2005.

R. Kirby, R. Simmons, and J. Forlizzi. Companion: A constraint-optimizing method for person-acceptable navigation. In *Robot and Human Interactive Communication, 2009. RO-MAN 2009. The 18th IEEE International Symposium on*, pages 607–612, Sept 2009.

N. Kirchner, A. Alempijevic, and G. Dissanayake. Nonverbal robot-group interaction using an imitated gaze cue. In *Proceedings of the 6th international conference on Human-robot interaction*, pages 497–504. ACM, 2011.

N. H. Kirk, D. Nyga, and M. Beetz. Controlled natural languages for language generation in artificial cognition. In *Robotics and Automation (ICRA), 2014 IEEE International Conference on*, pages 6667–6672. IEEE, 2014.

T. Kishi, T. Kojima, N. Endo, M. Destephe, T. Otani, L. Jamone, P. Kryczka, G. Trovato, K. Hashimoto, S. Cosentino, et al. Impression survey of the emotion expression humanoid robot with mental model based dynamic emotions. In *Robotics and Automation (ICRA), 2013 IEEE International Conference on*, pages 1663–1668. IEEE, 2013.

T. Kitade, S. Satake, T. Kanda, and M. Imai. Understanding suitable locations for waiting. In *Proceedings of the 8th ACM/IEEE International Conference on Human-robot Interaction*, HRI '13, pages 57–64, Piscataway, NJ, USA, 2013. IEEE Press. ISBN 978-1-4673-3055-8. URL `http://dl.acm.org.ezprimo1.idc.ac.il/citation.cfm?id=2447556.2447566`.

M. Knapp, J. Hall, and T. Horgan. *Nonverbal communication in human interaction*. Cengage Learning, 2013.

R. A. Knepper and D. Rus. Pedestrian-inspired sampling-based multi-robot collision avoidance. In *RO-MAN, 2012 IEEE*, pages 94–100, Sept 2012.

R. A. Knepper, S. Tellex, A. Li, N. Roy, and D. Rus. Recovering from failure by asking for help. *Autonomous Robots*, 39(3):347–362, 2015.

H. Knight and R. Simmons. Expressive motion with x, y and theta: Laban effort features for mobile robots. In *Robot and Human Interactive Communication, 2014 RO-MAN: The 23rd IEEE International Symposium on*, pages 267–273. IEEE, 2014.

H. Knight, R. Toscano, W. D. Stiehl, A. Chang, Y. Wang, and C. Breazeal. Real-time social touch gesture recognition for sensate robots. In *Intelligent Robots and Systems, 2009. IROS 2009. IEEE/RSJ International Conference on*, pages 3715–3720. IEEE, 2009.

W. B. Knox, P. Stone, and C. Breazeal. Training a robot via human feedback: A case study. In *Social Robotics*, pages 460–470. Springer, 2013.

T. Kollar, S. Tellex, D. Roy, and N. Roy. Toward understanding natural language directions. In *Human-Robot Interaction (HRI), 2010 5th ACM/IEEE International Conference on*, pages 259–266. IEEE, 2010.

T. Kollar, V. Perera, D. Nardi, and M. Veloso. Learning environmental knowledge from task-based human-robot dialog. In *Robotics and Automation (ICRA), 2013 IEEE International Conference on*, pages 4304–4309, May 2013.

Y. Kondo, K. Takemura, J. Takamatsu, and T. Ogasawara. A gesture-centric android system for multi-party human-robot interaction. *Journal of Human-Robot Interaction*, 2(1):133–151, 2013.

S. Koo and D.-S. Kwon. Recognizing human intentional actions from the relative movements between human and robot. In *Robot and Human Interactive Communication, 2009. RO-MAN 2009. The 18th IEEE International Symposium on*, pages 939–944, Sept 2009.

H. S. Koppula and A. Saxena. Anticipating human activities using object affordances for reactive robotic response. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 38(1):14–29, 2016.

H. S. Koppula, R. Gupta, and A. Saxena. Learning human activities and object affordances from rgb-d videos. *The International Journal of Robotics Research*, 32(8):951–970, 2013.

T. Kruse, A. Kirsch, H. Khambhaita, and R. Alami. Evaluating directional cost models in navigation. In *Proceedings of the 2014 ACM/IEEE International Conference on Human-robot Interaction*, HRI '14, pages 350–357, New York, NY, USA, 2014. ACM. ISBN 978-1-4503-2658-2. URL http://doi.acm.org.ezprimo1.idc.ac.il/10.1145/2559636.2559662.

M. Kuderer and W. Burgard. An approach to socially compliant leader following for mobile robots. In *Social Robotics*, pages 239–248. Springer, 2014.

D. Kulic and E. A. Croft. Affective state estimation for human–robot interaction. *Robotics, IEEE Transactions on*, 23(5):991–1000, 2007.

Y. Kuno, H. Sekiguchi, T. Tsubota, S. Moriyama, K. Yamazaki, and A. Yamazaki. Museum guide robot with communicative head motion. In *Robot and Human Interactive Communication, 2006. ROMAN 2006. The 15th IEEE International Symposium on*, pages 33–38. IEEE, 2006.

W. Y. Kwon and I. H. Suh. A temporal bayesian network with application to design of a proactive robotic assistant. In *Robotics and Automation (ICRA), 2012 IEEE International Conference on*, pages 3685–3690. IEEE, 2012.

R. Laban and L. Ullmann. *The mastery of movement.* ERIC, 1971.

C. Lang, S. Wachsmuth, M. Hanheide, and H. Wersing. Facial communicative signal interpretation in human-robot interaction by discriminative video subsequence selection. In *Robotics and Automation (ICRA), 2013 IEEE International Conference on*, pages 170–177. IEEE, 2013.

J. Lasseter. Principles of traditional animation applied to 3d computer animation. In *SIGGRAPH '87: Proceedings of the 14th annual conference on Computer graphics and interactive techniques*, pages 35–44, New York, NY, USA, 1987. ACM.

B. Lau, K. O. Arras, and W. Burgard. Multi-model hypothesis group tracking and group size estimation. *International Journal of Social Robotics*, 2(1): 19–30, 2010.

J. Lee, J. F. Kiser, A. F. Bobick, and A. L. Thomaz. Vision-based contingency detection. In *Proceedings of the 6th international conference on Human-robot interaction*, pages 297–304. ACM, 2011.

J. Lee, C. Chao, A. F. Bobick, and A. L. Thomaz. Multi-cue contingency detection. *International Journal of Social Robotics*, 4(2):147–161, 2012.

M. K. Lee, S. Kiesler, J. Forlizzi, S. Srinivasa, and P. Rybski. Gracefully mitigating breakdowns in robotic services. In *Human-Robot Interaction (HRI), 2010 5th ACM/IEEE International Conference on*, pages 203–210. IEEE, 2010.

I. Leite, R. Henriques, C. Martinho, and A. Paiva. Sensors in the wild: Exploring electrodermal activity in child-robot interaction. In *Proceedings of the 8th ACM/IEEE international conference on Human-robot interaction*, pages 41–48. IEEE Press, 2013.

S. Lemaignan, R. Ros, E. A. Sisbot, R. Alami, and M. Beetz. Grounding the interaction: Anchoring situated discourse in everyday human-robot interaction. *International Journal of Social Robotics*, 4(2):181–199, 2012.

C. Lenz, S. Nair, M. Rickert, A. Knoll, W. Rosel, J. Gast, A. Bannat, and F. Wallhoff. Joint-action for humans and industrial robots for assembly tasks. In *Robot and Human Interactive Communication, 2008. RO-MAN 2008. The 17th IEEE International Symposium on*, pages 130–135. IEEE, 2008.

Y. Li, K. P. Tee, W. L. Chan, R. Yan, Y. Chua, and D. K. Limbu. Continuous role adaptation for human–robot shared control. *Robotics, IEEE Transactions on*, 31(3):672–681, 2015.

C. Lichtenthäler, A. Peters, S. Griffiths, and A. Kirsch. Social navigation-identifying robot navigation patterns in a path crossing scenario. In *Social Robotics*, pages 84–93. Springer, 2013.

C. Liu, K. Conn, N. Sarkar, and W. Stone. Online affect detection and robot behavior adaptation for intervention of children with autism. *Robotics, IEEE Transactions on*, 24(4):883–896, 2008.

P. Liu, D. F. Glas, T. Kanda, H. Ishiguro, and N. Hagita. It's not polite to point: generating socially-appropriate deictic behaviors towards people. In *Proceedings of the 8th ACM/IEEE international conference on Human-robot interaction*, pages 267–274. IEEE Press, 2013.

A. Lockerd and C. Breazeal. Tutelage and socially guided robot learning. In *Intelligent Robots and Systems, 2004.(IROS 2004). Proceedings. 2004 IEEE/RSJ International Conference on*, volume 4, pages 3475–3480. IEEE, 2004.

M. Lohse, K. J. Rohlfing, B. Wrede, and G. Sagerer. Try something else! when users change their discursive behavior in human-robot interaction. In *Robotics and Automation, 2008. ICRA 2008. IEEE International Conference on*, pages 3481–3486. IEEE, 2008.

M. Lohse, R. Rothuis, J. Gallego-Pérez, D. E. Karreman, and V. Evers. Robot gestures make difficult tasks easier: the impact of gestures on perceived workload and task performance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1459–1466. ACM, 2014.

M. Luber and K. O. Arras. Multi-hypothesis social grouping and tracking for mobile robots. In *Robotics: Science and Systems*, 2013.

M. Luber, L. Spinello, J. Silva, and K. O. Arras. Socially-aware robot navigation: A learning approach. In *Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on*, pages 902–907, Oct 2012.

I. Lütkebohle, F. Hegel, S. Schulz, M. Hackel, B. Wrede, S. Wachsmuth, and G. Sagerer. The bielefeld anthropomorphic robot head "flobi". In *Robotics and Automation, 2010. ICRA 2010. Proceedings IEEE International Conference on*, volume 3(7), pages 3384–3391, 2010.

J. MacGlashan, M. Babes-Vroman, M. desJardins, M. Littman, S. Muresan, S. Squire, S. Tellex, D. Arumugam, and L. Yang. Grounding english commands to reward functions. In *Robotics: Science and Systems*, 2015.

J. Mainprice and D. Berenson. Human-robot collaborative manipulation planning using early prediction of human motion. In *Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on*, pages 299–306. IEEE, 2013.

J. Mainprice, E. A. Sisbot, L. Jaillet, J. Cortés, R. Alami, and T. Siméon. Planning human-aware motions using a sampling-based costmap planner. In *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, pages 5012–5017. IEEE, 2011.

J. Mainprice, M. Gharbi, T. Siméon, and R. Alami. Sharing effort in planning human-robot handover tasks. In *RO-MAN, 2012 IEEE*, pages 764–770. IEEE, 2012.

M. Mason and M. Lopes. Robot self-initiative and personalization by learning through repeated interactions. In *Human-Robot Interaction (HRI), 2011 6th ACM/IEEE International Conference on*, pages 433–440. IEEE, 2011.

M. Masuda and S. Kato. Motion rendering system for emotion expression of human form robots based on laban movement analysis. In *RO-MAN, 2010 IEEE*, pages 324–329. IEEE, 2010.

T. Matsumaru, K. Iwase, K. Akiyama, T. Kusada, and T. Ito. Mobile robot with eyeball expression as the preliminary-announcement and display of the robot's following motion. *Autonomous Robots*, 18(2):231–246, 2005.

C. Matuszek, D. Fox, and K. Koscher. Following directions using statistical machine translation. In *Proceedings of the 5th ACM/IEEE international conference on Human-robot interaction*, pages 251–258. IEEE Press, 2010.

B. A. Maxwell. Building robot systems to interact with people in real environments. *Autonomous Robots*, 22(4):353–367, 2007.

D. McColl and G. Nejat. Affect detection from body language during social hri. In *RO-MAN, 2012 IEEE*, pages 1013–1018. IEEE, 2012.

D. McColl, Z. Zhang, and G. Nejat. Human body pose interpretation and classification for social human-robot interaction. *International Journal of Social Robotics*, 3(3):313–332, 2011.

S. McKeague, J. Liu, and G.-Z. Yang. An asynchronous rgb-d sensor fusion framework using monte-carlo methods for hand tracking on a mobile robot in crowded environments. In *Social Robotics*, pages 491–500. Springer, 2013.

R. Mead, A. Atrash, and M. J. Matarić. Proxemic feature recognition for interactive robots: automating metrics from the social sciences. In *Social Robotics*, pages 52–61. Springer, 2011.

R. Mead, A. Atrash, and M. J. Matarić. Automated proxemic feature extraction and behavior recognition: Applications in human-robot interaction. *International Journal of Social Robotics*, 5(3):367–378, 2013.

J. R. Medina, T. Lorenz, and S. Hirche. Synthesizing anticipatory haptic assistance considering human behavior uncertainty. *Robotics, IEEE Transactions on*, 31(1):180–190, 2015.

E. Meisner, V. Isler, and J. Trinkle. Controller design for human-robot interaction. *Autonomous Robots*, 24(2):123–134, 2008.

A. N. Meltzoff. The human infant as imitative generalist: A 20-year progress report on infant imitation with implications for comparative psychology. In B. G. Galef C. M. Heyes, editor, *Social Learning in Animals: The Roots of Culture*. Academic Press, San Diego, CA, 1996.

Ç. Meriçli, M. Veloso, and H. L. Akın. Multi-resolution corrective demonstration for efficient task execution and refinement. *International Journal of Social Robotics*, 4(4):423–435, 2012.

M. P. Michalowski and R. Simmons. Multimodal person tracking and attention classification. In *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*, pages 347–348. ACM, 2006.

M. P. Michalowski, S. Sabanovic, and H. Kozima. A dancing robot for rhythmic social interaction. In *Human-Robot Interaction (HRI), 2007 2nd ACM/IEEE International Conference on*, pages 89–96. IEEE, 2007.

F. Michaud, C. Côté, D. Létourneau, Y. Brosseau, J-M. Valin, É. Beaudry, C. Raïevsky, A. Ponchon, P. Moisan, P. Lepage, et al. Spartacus attending the 2005 AAAI conference. *Autonomous Robots*, 22(4):369–383, 2007.

T. Miller, A. Exley, and W. Schuler. Elements of a spoken language programming interface for robots. In *Proceedings of the ACM/IEEE international conference on Human-robot interaction*, pages 231–237. ACM, 2007.

G. Milliez, M. Warnier, A. Clodic, and R. Alami. A framework for endowing an interactive robot with reasoning capabilities about perspective-taking and belief management. In *Robot and Human Interactive Communication, 2014 RO-MAN: The 23rd IEEE International Symposium on*, pages 1103–1109. IEEE, 2014.

T. Minato and H. Ishiguro. Construction and evaluation of a model of natural human motion based on motion diversity. In *Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction*, pages 65–72. ACM, 2008.

M. Minsky. Music, mind, and meaning. In *Music, mind, and brain*, pages 1–19. Springer, 1982.

Y. Mohammad and T. Nishida. Fluid imitation. *International Journal of Social Robotics*, 4(4):369–382, 2012.

A. Moon, D. M. Troniak, B. Gleeson, M. K. X. J. Pan, M. Zheng, B. A. Blumer, K. MacLean, and E. A. Croft. Meet me where i'm gazing: how shared attention gaze affects human-robot handover timing. In *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*, pages 334–341. ACM, 2014.

N.-J. Moore, M. Hickson, and D. W. Stacks. *Nonverbal communication: Studies and applications*. Oxford University Press, 6th ed edition, 2013.

S. Morales, Y. Luis, S. Satake, R. Huq, D. Glas, T. Kanda, and N. Hagita. How do people walk side-by-side?: Using a computational model of human behavior for a social robot. In *Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction*, HRI '12, pages 301–308, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1063-5. URL `http://doi.acm.org.ezprimo1.idc.ac.il/10.1145/2157689.2157799`.

A. Mörtl, M. Lawitzky, A. Kucukyilmaz, M. Sezgin, C. Basdogan, and S. Hirche. The role of roles: Physical cooperation between humans and robots. *The International Journal of Robotics Research*, 31(13):1656–1674, 2012.

L. Moshkina and R. C. Arkin. Human perspective on affective robotic behavior: A longitudinal study. In *Intelligent Robots and Systems, 2005.(IROS 2005). 2005 IEEE/RSJ International Conference on*, pages 1444–1451. IEEE, 2005.

Q. Muhlbauer, S. Sosnowski, X. Tingting, Z. Tianguang, K. Kuhnlenz, and M. Buss. Navigation through urban environments by visual perception and interaction. In *Robotics and Automation, 2009. ICRA '09. IEEE International Conference on*, pages 3558–3564, May 2009.

M. Mühlig, M. Gienger, and J. J Steil. Interactive imitation learning of object movement skills. *Autonomous Robots*, 32(2):97–114, 2012.

R. Murakami, L. Y. Morales Saiki, S. Satake, T. Kanda, and H. Ishiguro. Destination unknown: Walking side-by-side without knowing the goal. In *Proceedings of the 2014 ACM/IEEE International Conference on Human-robot Interaction*, HRI '14, pages 471–478, New York, NY, USA, 2014. ACM. ISBN 978-1-4503-2658-2. URL `http://doi.acm.org.ezprimo1.idc.ac.il/10.1145/2559636.2559665`.

B. Mutlu, T. Shiwa, T. Kanda, H. Ishiguro, and N. Hagita. Footing in human-robot conversations: how robots might shape participant roles using gaze cues. In *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction*, pages 61–68. ACM, 2009a.

B. Mutlu, F. Yamaoka, T. Kanda, H. Ishiguro, and N. Hagita. Nonverbal leakage in robots: communication of intentions through seemingly unintentional behavior. In *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction*, pages 69–76. ACM, 2009b.

Y. Nagai. The role of motion information in learning human-robot joint attention. In *Robotics and Automation, 2005. ICRA 2005. Proceedings of the 2005 IEEE International Conference on*, pages 2069–2074. IEEE, 2005.

Y. Nagai, C. Muhl, and K. J. Rohlfing. Toward designing a robot that learns actions from parental demonstrations. In *Robotics and Automation, 2008. ICRA 2008. IEEE International Conference on*, pages 3545–3550. IEEE, 2008.

N. Najmaei and M. R. Kermani. Prediction-based reactive control strategy for human-robot interactions. In *Robotics and Automation (ICRA), 2010 IEEE International Conference on*, pages 3434–3439. IEEE, 2010.

K. Nakagawa, M. Shiomi, K. Shinozawa, R. Matsumura, H. Ishiguro, and N. Hagita. Effect of robot's active touch on people's motivation. In *Proceedings of the 6th international conference on Human-robot interaction*, pages 465–472. ACM, 2011.

K. Nakagawa, M. Shiomi, K. Shinozawa, R. Matsumura, H. Ishiguro, and N. Hagita. Effect of robotâăŹ™s whispering behavior on peopleâăŹ™s motivation. *International Journal of Social Robotics*, 5(1):5–16, 2013.

K. Namera, S. Takasugi, K. Takano, T. Yamamoto, and Y. Miyake. Timing control of utterance and body motion in human-robot interaction. In *Robot and Human Interactive Communication, 2008. RO-MAN 2008. The 17th IEEE International Symposium on*, pages 119–123. IEEE, 2008.

L. Nardi and L. Iocchi. Representation and execution of social plans through human-robot collaboration. In *Social Robotics*, pages 266–275. Springer, 2014.

A. Niculescu, B. van Dijk, A. Nijholt, H. Li, and S. L. See. Making social robots more attractive: the effects of voice pitch, humor and empathy. *International journal of social robotics*, 5(2):171–191, 2013.

S. Nikolaidis and J. Shah. Human-robot cross-training: computational formulation, modeling and evaluation of a human team training strategy. In *Proceedings of the 8th ACM/IEEE international conference on Human-robot interaction*, pages 33–40. IEEE Press, 2013.

S. Nikolaidis, R. Ramakrishnan, K. Gu, and J. Shah. Efficient model learning from joint-action demonstrations for human-robot collaborative tasks. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, pages 189–196. ACM, 2015.

D. Norman. *Emotional Design: Why We Love (or Hate) Everyday Things.* Basic Books, New York, 2004.

D. Nyga, M. Tenorth, and M. Beetz. How-models of human reaching movements in the context of everyday manipulation activities. In *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, pages 6221–6226. IEEE, 2011.

Y. Okuno, T. Kanda, M. Imai, H. Ishiguro, and N. Hagita. Providing route directions: design of robot's utterance, gesture, and timing. In *Human-Robot Interaction (HRI), 2009 4th ACM/IEEE International Conference on*, pages 53–60. IEEE, 2009.

S. Ou and R. Grupen. From manipulation to communicative gesture. In *Proceedings of the 5th ACM/IEEE international conference on Human-robot interaction*, pages 325–332. IEEE Press, 2010.

E. Pacchierotti, H. I. Christensen, and P. Jensfelt. Evaluation of passing distance for social robots. In *Robot and Human Interactive Communication, 2006. ROMAN 2006. The 15th IEEE International Symposium on*, pages 315–320, Sept 2006.

A. Panangadan, M. Matarić, and G. S. Sukhatme. Tracking and modeling of human activity using laser rangefinders. *International Journal of Social Robotics*, 2(1):95–107, 2010.

A. K. Pandey and R. Alami. A framework towards a socially aware mobile robot motion in human-centered dynamic environment. In *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*, pages 5855–5860, Oct 2010a.

A. K. Pandey and R. Alami. Mightability maps: A perceptual level decisional framework for co-operative and competitive human-robot interaction. In *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*, pages 5842–5848. IEEE, 2010b.

P. Papadakis, P. Rives, and A. Spalanzani. Adaptive spacing in human-robot interactions. In *Intelligent Robots and Systems (IROS 2014), 2014 IEEE/RSJ International Conference on*, pages 2627–2632, Sept 2014.

J.-C. Park, H.-R. Kim, Y.-M. Kim, and D.-S. Kwon. Robot's individual emotion generation model and action coloring according to the robot's personality. In *Robot and Human Interactive Communication, 2009. RO-MAN 2009. The 18th IEEE International Symposium on*, pages 257–262. IEEE, 2009.

M. Pateraki, H. Baltzakis, P. Kondaxakis, and P. Trahanias. Tracking of facial features to support human-robot interaction. In *Robotics and Automation, 2009. ICRA'09. IEEE International Conference on*, pages 3755–3760. IEEE, 2009.

C. Perez Quintero, R. T. Fomena, A. Shademan, N. Wolleb, T. Dick, and M. Jagersand. Sepo: Selecting by pointing as an intuitive human-robot command interface. In *Robotics and Automation (ICRA), 2013 IEEE International Conference on*, pages 1166–1171. IEEE, 2013.

L. Peternel, T. Petrič, E. Oztop, and J. Babič. Teaching robots to cooperate with humans in dynamic manipulation tasks based on multi-modal human-in-the-loop approach. *Autonomous robots*, 36(1-2):123–136, 2014.

R. W. Picard. *Affective computing*, volume 252. MIT press Cambridge, 1997.

B. Raducanu and F. Dornaika. Dynamic facial expression recognition using laplacian eigenmaps-based manifold learning. In *Robotics and Automation (ICRA), 2010 IEEE International Conference on*, pages 156–161. IEEE, 2010.

M. Ralph and M. A. Moussa. Toward a natural language interface for transferring grasping skills to robots. *Robotics, IEEE Transactions on*, 24(2): 468–475, 2008.

V. Raman, C. Lignos, C. Finucane, K. C. T. Lee, M. P. Marcus, and H. Kress-Gazit. Sorry dave, i'm afraid i can't do that: Explaining unachievable robot tasks using natural language. In *Robotics: Science and Systems*, volume 2. Citeseer, 2013.

P. Ratsamee, Y. Mae, K. Ohara, M. Kojima, and T. Arai. Social navigation model based on human intention analysis using face orientation. In *Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on*, pages 1682–1687, Nov 2013.

R. Read and T. Belpaeme. Situational context directs how people affectively interpret robotic non-linguistic utterances. In *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*, pages 41–48. ACM, 2014.

T. Ribeiro and A. Paiva. The illusion of robotic life: principles and practices of animation for robots. In *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*, pages 383–390. ACM, 2012.

C. Rich, B. Ponsler, A. Holroyd, and C. L. Sidner. Recognizing engagement in human-robot interaction. In *Human-Robot Interaction (HRI), 2010 5th ACM/IEEE International Conference on*, pages 375–382. IEEE, 2010.

L. D. Riek, T.-C. Rabinowitch, P. Bremner, A. G. Pipe, M. Fraser, and P. Robinson. Cooperative gestures: Effective signaling for humanoid robots. In *Human-Robot Interaction (HRI), 2010 5th ACM/IEEE International Conference on*, pages 61–68. IEEE, 2010.

J. Rios-Martinez, A. Spalanzani, and C. Laugier. From proxemics theory to socially-aware navigation: A survey. *International Journal of Social Robotics*, 7(2):137–153, 2015.

R. Ros, S. Lemaignan, E. A. Sisbot, R. Alami, J. Steinwender, K. Hamann, and F. Warneken. Which one? grounding the referent based on efficient human-robot interaction. In *RO-MAN, 2010 IEEE*, pages 570–575. IEEE, 2010.

J. A. Russell. Core affect and the psychological construction of emotion. *Psychological review*, 110(1):145, 2003.

P. E. Rybski, K. Yoon, J. Stolarz, and M. M. Veloso. Interactive robot task training through dialog and demonstration. In *Human-Robot Interaction (HRI), 2007 2nd ACM/IEEE International Conference on*, pages 49–56. IEEE, 2007.

M. S. Ryoo, T. J. Fuchs, L. Xia, J. K. Aggarwal, and L. Matthies. Robot-centric activity prediction from first-person videos: What will they do to me'. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, pages 295–302. ACM, 2015.

J. Saldien, K. Goris, B. Vanderborght, J. Vanderfaeillie, and D. Lefeber. Expressing emotions with the social robot probo. *International Journal of Social Robotics*, 2(4):377–389, 2010.

M. Salem, S. Kopp, I. Wachsmuth, K. Rohlfing, and F. Joublin. Generation and evaluation of communicative robot gesture. *International Journal of Social Robotics*, 4(2):201–217, 2012.

M. Salem, S. Kopp, and F. Joublin. Closing the loop: Towards tightly synchronized robot gesture and speech. In *Social Robotics*, pages 381–391. Springer, 2013a.

M. Salem, M. Ziadee, and M. Sakr. Effects of politeness and interaction context on perception and experience of hri. In *Social Robotics*, pages 531–541. Springer, 2013b.

M. Salem, M. Ziadee, and M. Sakr. Marhaba, how may i help you?: Effects of politeness and culture on robot acceptance and anthropomorphization. In *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*, pages 74–81. ACM, 2014.

J. Sanghvi, G. Castellano, I. Leite, A. Pereira, P. W. McOwan, and A. Paiva. Automatic analysis of affective postures and body motion to detect engagement with a game companion. In *Human-Robot Interaction (HRI), 2011 6th ACM/IEEE International Conference on*, pages 305–311. IEEE, 2011.

S. Satake, T. Kanda, D. F. Glas, M. Imai, H. Ishiguro, and N. Hagita. A robot that approaches pedestrians. *Robotics, IEEE Transactions on*, 29(2): 508–524, April 2013. ISSN 1552-3098.

S. Satake, H. Iba, T. Kanda, M. Imai, and Y. M. Saiki. May i talk about other shops here?: Modeling territory and invasion in front of shops. In *Proceedings of the 2014 ACM/IEEE International Conference on Human-robot Interaction*, HRI '14, pages 487–494, New York, NY, USA, 2014. ACM. ISBN 978-1-4503-2658-2. URL `http://doi.acm.org.ezprimo1.idc.ac.il/10.1145/2559636.2559669`.

J. Sattar and G. Dudek. Towards quantitative modeling of task confirmations in human-robot dialog. In *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, pages 1957–1963. IEEE, 2011.

J. Sattar and J. J. Little. Ensuring safety in human-robot dialog—a cost-directed approach. In *Robotics and Automation (ICRA), 2014 IEEE International Conference on*, pages 6660–6666. IEEE, 2014.

A. Sauppé and B. Mutlu. Robot deictics: How gesture and context shape referential communication. In *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*, pages 342–349. ACM, 2014.

B. Scassellati. Foundations for a theory of mind for a humanoid robot. *Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science, PhD Thesis*, 2001.

K. R. Scherer. What are emotions? and how can they be measured? *Social science information*, 44(4):695–729, 2005.

A. J. Schmid, O. Weede, and H. Wörn. Proactive robot task selection given a human intention estimate. In *Robot and Human interactive Communication, 2007. RO-MAN 2007. The 16th IEEE International Symposium on*, pages 726–731. IEEE, 2007.

O. C. Schrempf, U. D. Hanebeck, A. J. Schmid, and H. Worn. A novel approach to proactive human-robot cooperation. In *Robot and Human Interactive Communication, 2005. ROMAN 2005. IEEE International Workshop on*, pages 555–560. IEEE, 2005.

N. Sebanz, H. Bekkering, and G. Knoblicha. Joint action: Bodies and minds moving together. *Trends in Cognitive Science*, 2006.

J. Shah, J. Wiken, B. Williams, and C. Breazeal. Improved human-robot team performance using chaski, a human-inspired plan execution system. In *Proceedings of the 6th international conference on Human-robot interaction*, pages 29–36. ACM, 2011.

M. Sharma, D. Hildebrandt, G. Newman, J. E. Young, and R. Eskicioglu. Communicating affect via flight path exploring use of the laban effort system for designing affective locomotion paths. In *Human-Robot Interaction (HRI), 2013 8th ACM/IEEE International Conference on*, pages 293–300. IEEE, 2013.

C. Shi, T. Kanda, M. Shimada, F. Yamaoka, H. Ishiguro, and N. Hagita. Easy development of communicative behaviors in social robots. In *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*, pages 5302–5309. IEEE, 2010.

H. Shibata, M. Kano, S. Kato, and H. Itoh. A system for converting robot'emotion'into facial expressions. In *Robotics and Automation, 2006. ICRA 2006. Proceedings 2006 IEEE International Conference on*, pages 3660–3665. IEEE, 2006.

M. Shiomi, T. Kanda, H. Ishiguro, and N. Hagita. A larger audience, please!: encouraging people to listen to a guide robot. In *Proceedings of the 5th ACM/IEEE international conference on Human-robot interaction*, pages 31–38. IEEE Press, 2010.

M. Shiomi, F. Zanlungo, K. Hayashi, and T. Kanda. Towards a socially acceptable collision avoidance for a mobile robot navigating among pedestrians using a pedestrian model. *International Journal of Social Robotics*, 6(3): 443–455, 2014.

C. L. Sidner, C. Lee, L.-P. Morency, and C. Forlines. The effect of head-nod recognition in human-robot conversation. In *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*, pages 290–296. ACM, 2006.

D. Silvera-Tawil, D. Rye, and M. Velonaki. Interpretation of social touch on an artificial arm covered with an eit-based sensitive skin. *International Journal of Social Robotics*, 6(4):489–505, 2014.

E. A. Sisbot and R. Alami. A human-aware manipulation planner. *Robotics, IEEE Transactions on*, 28(5):1045–1057, 2012.

E. A. Sisbot, L. F. Marin-Urias, R. Alami, and T. Simeon. A human aware mobile robot motion planner. *Robotics, IEEE Transactions on*, 23(5):874–883, Oct 2007. ISSN 1552-3098.

E. A. Sisbot, L. F. Marin-Urias, X. Broquere, D. Sidobre, and R. Alami. Synthesizing robot motions adapted to human presence. *International Journal of Social Robotics*, 2(3):329–343, 2010.

M. Sorostinean, F. Ferland, A. Tapus, et al. Motion-oriented attention for a social gaze robot behavior. In *Social Robotics*, pages 310–319. Springer, 2014.

A. St Clair and M. J. Mataric. How robot verbal feedback can improve team performance in human-robot task collaborations. In *HRI*, pages 213–220, 2015.

C. Stanton and C. J. Stevens. Robot pressure: the impact of robot eye gaze and lifelike bodily movements upon decision-making and trust. In *Social Robotics*, pages 330–339. Springer, 2014.

M. Staudte and M. W. Crocker. Visual attention in spoken human-robot interaction. In *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction*, pages 77–84. ACM, 2009.

W. D. Stiehl, J. Lieberman, C. Breazeal, L. Basel, L. Lalla, and M. Wolf. Design of a therapeutic robotic companion for relational, affective touch. In *Robot and Human Interactive Communication, 2005. ROMAN 2005. IEEE International Workshop on*, pages 408–415. IEEE, 2005.

K. W. Strabala, M. K. Lee, A. D. Dragan, J. L. Forlizzi, S. Srinivasa, M. Cakmak, and V. Micelli. Towards seamless human-robot handovers. *Journal of Human-Robot Interaction*, 2(1):112–132, 2013.

H. B. Suay and S. Chernova. Effect of human guidance and state space size on interactive reinforcement learning. In *RO-MAN, 2011 IEEE*, pages 1–6. IEEE, 2011.

H. B. Suay, R. Toris, and S. Chernova. A practical comparison of three robot learning from demonstration algorithm. *International Journal of Social Robotics*, 4(4):319–330, 2012.

K. Sugiura, Y. Shiga, H. Kawai, T. Misu, and C. Hori. Non-monologue hmm-based speech synthesis for service robots: A cloud robotics approach. In *Robotics and Automation (ICRA), 2014 IEEE International Conference on*, pages 2237–2242. IEEE, 2014.

M. Svenstrup, S. Tranberg, H. J. Andersen, and T. Bak. Pose estimation and adaptive robot behaviour for human-robot interaction. In *Robotics and Automation, 2009. ICRA'09. IEEE International Conference on*, pages 3571–3576. IEEE, 2009.

E. Sviestins, N. Mitsunaga, T. Kanda, H. Ishiguro, and N. Hagita. Speed adaptation for a robot walking with a human. In *Human-Robot Interaction (HRI), 2007 2nd ACM/IEEE International Conference on*, pages 349–356, March 2007.

D. Szafir, B. Mutlu, and T. Fong. Communication of intent in assistive free flyers. In *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*, pages 358–365. ACM, 2014.

D. Szafir, B. Mutlu, and T. Fong. Communicating directionality in flying robots. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, pages 19–26. ACM, 2015.

L. Takayama, D. Dooley, and W. Ju. Expressing thought: improving robot readability with animation principles. In *Proceedings of the 6th international conference on Human-robot interaction*, pages 69–76. ACM, 2011.

K. Talamadupula, G. Briggs, T. Chakraborti, M. Scheutz, and S. Kambhampati. Coordination in human-robot teams using mental modeling and plan recognition. In *Intelligent Robots and Systems (IROS 2014), 2014 IEEE/RSJ International Conference on*, pages 2957–2962. IEEE, 2014.

B. Tanya, C. Malinda, C. Josep, and T. Michael. Unwilling Versus Unable: Infants' Understanding of Intentional Action. *Developmental Psychology*, 41:328–337, 2005.

T. Tasaki, K. Komatani, T. Ogata, and H. G. Okuno. Spatially mapping of friendliness for human-robot interaction. In *Intelligent Robots and Systems, 2005.(IROS 2005). 2005 IEEE/RSJ International Conference on*, pages 1277–1282. IEEE, 2005.

S. Tellex, T. Kollar, S. Dickerson, M. R. Walter, A. G. Banerjee, S. Teller, and N. Roy. Approaching the symbol grounding problem with probabilistic graphical models. *AI magazine*, 32(4):64–76, 2011.

S. Tellex, P. Thaker, R. Deitsl, D. Simeonovl, T. Kollar, and N. Royl. Toward information theoretic human-robot dialog. *Robotics*, page 409, 2013.

S. Tellex, R. Knepper, A. Li, D. Rus, and N. Roy. Asking for help using inverse semantics. In *Robotics: Science and systems*, volume 2, page 3, 2014.

C. L. Teo, Y. Yang, H. Daumé III, C. Fermüller, and Y. Aloimonos. Towards a watson that sees: Language-guided action recognition for robots. In *Robotics and Automation (ICRA), 2012 IEEE International Conference on*, pages 374–381. IEEE, 2012.

F. Thomas and O. Johnston. *Disney Animation: The Illusion of Life.* Abbeville Press, New York, 1981.

A. L. Thomaz. *Socially Guided Machine Learning.* PhD thesis, Massachusetts Institute of Technology, 2006.

A. L. Thomaz and M. Cakmak. Learning about objects with human teachers. In *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction*, pages 15–22. ACM, 2009.

A. L. Thomaz, G. Hoffman, and C. Breazeal. Experiments in socially guided machine learning: understanding how humans teach. In *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*, pages 359–360. ACM, 2006a.

A. L. Thomaz, G. Hoffman, and C. Breazeal. Reinforcement learning with human teachers: Understanding how people want to teach robots. In *Robot and Human Interactive Communication, 2006. ROMAN 2006. The 15th IEEE International Symposium on*, pages 352–357. IEEE, 2006b.

M. Tomasello. *The Cultural Origins of Human Cognition*. Harvard University Press, March 2001. ISBN 0674005821.

M. Tomasello, M. Carptenter, J. Call, T. Behne, and H. Moll. Understanding and sharing intentions: the origins of cultural cognition. *Behavioral and Brain Sciences (in press)*, 2004.

E. A. Topp and H. I. Christensen. Detecting region transitions for human-augmented mapping. *Robotics, IEEE Transactions on*, 26(4):715–720, Aug 2010. ISSN 1552-3098.

C. Torrey, A. Powers, S. R. Fussell, and S. Kiesler. Exploring adaptive dialogue based on a robot's awareness of human gaze and task progress. In *Proceedings of the ACM/IEEE international conference on Human-robot interaction*, pages 247–254. ACM, 2007.

C. Torrey, S. Fussell, and S. Kiesler. How a robot should give advice. In *Proceedings of the 8th ACM/IEEE international conference on Human-robot interaction*, pages 275–282. IEEE Press, 2013.

E. Torta, R. H. Cuijpers, J. F. Juola, and D. van der Pol. Design of robust robotic proxemic behaviour. In *Social robotics*, pages 21–30. Springer, 2011.

E. Torta, J. van Heumen, R. H. Cuijpers, and J. F. Juola. How can a robot attract the attention of its human partner? a comparative study over different modalities for attracting attention. In *Social robotics*, pages 288–297. Springer, 2012.

E. Torta, J. van Heumen, F. Piunti, L. Romeo, and R. Cuijpers. Evaluation of unimodal and multimodal communication cues for attracting attention in human–robot interaction. *International Journal of Social Robotics*, 7(1): 89–96, 2015.

G. Trafton, L. Hiatt, A. Harrison, F. Tamborello, S. Khemlani, and A. Schultz. ACT-R/E: An embodied cognitive architecture for human-robot interaction. *Journal of Human-Robot Interaction*, 2(1):30–55, 2013.

J. G. Trafton, M. D. Bugajska, B. R. Fransen, and R. M. Ratwani. Integrating vision and audition within a cognitive architecture to track conversations. In *Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction*, pages 201–208. ACM, 2008.

H. S. Tranberg, M. Svenstrup, H. J. Andersen, and T. Bak. Adaptive human aware navigation based on motion pattern analysis. In *Robot and Human Interactive Communication, 2009. RO-MAN 2009. The 18th IEEE International Symposium on*, pages 927–932, Sept 2009.

P. Trautman, J. Ma, R. M. Murray, and A. Krause. Robot navigation in dense human crowds: Statistical models and experimental studies of human–robot cooperation. *The International Journal of Robotics Research*, 34(3):335–356, 2015. URL `http://ijr.sagepub.com/content/34/3/335.abstract`.

J.-M. Valin, S. Yamamoto, J. Rouat, F. Michaud, K. Nakadai, and H. G. Okuno. Robust recognition of simultaneous speech by a mobile robot. *Robotics, IEEE Transactions on*, 23(4):742–752, 2007.

M. Van den Bergh, D. Carton, R. De Nijs, N. Mitsou, C. Landsiedel, K. Kuehnlenz, D. Wollherr, L. Van Gool, and M. Buss. Real-time 3d hand gesture interaction with a robot for understanding directions from humans. In *RO-MAN, 2011 IEEE*, pages 357–362. IEEE, 2011.

E. T. Van Dijk, E. Torta, and R. H. Cuijpers. Effects of eye contact and iconic gestures on message retention in human-robot interaction. *International Journal of Social Robotics*, 5(4):491–501, 2013.

G. Venture, H. Kadone, T. Zhang, J. Grèzes, A. Berthoz, and H. Hicheur. Recognizing emotions conveyed by human gait. *International Journal of Social Robotics*, 6(4):621–632, 2014.

L. S. Vygotsky and Ed. M. Cole. *Mind in society: the development of higher psychological processes.* Harvard University Press, Cambridge, MA, 1978.

Z. Wang, K. Mülling, M. P. Deisenroth, H. B. Amor, D. Vogt, B. Schölkopf, and J. Peters. Probabilistic movement modeling for intention inference in human–robot interaction. *The International Journal of Robotics Research*, 32(7):841–858, 2013.

A. Weiss, J. Igelsböck, M. Tscheligi, A. Bauer, K. Kühnlenz, D. Wollherr, and M. Buss. Robots asking for directions: the willingness of passers-by to support robots. In *Proceedings of the 5th ACM/IEEE international conference on Human-robot interaction*, pages 23–30. IEEE Press, 2010.

K. Williams and C. Breazeal. A reasoning architecture for human-robot joint tasks using physics-, social-, and capability-based logic. In *Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on*, pages 664–671. IEEE, 2012.

M.-A. Williams, S. Abidi, P. Gärdenfors, X. Wang, B. Kuipers, and B. Johnston. Interpreting robot pointing behavior. In *Proceedings of the 5th International Conference on Social Robotics-Volume 8239*, pages 148–159, 2013.

A. Xu and G. Dudek. Trust-driven interactive visual navigation for autonomous robots. In *Robotics and Automation (ICRA), 2012 IEEE International Conference on*, pages 3922–3929. IEEE, 2012.

T. Yamaguchi and S. Hashimoto. Attractiveeye: Augmented gaze representation for "what is the robot looking at?". In *Robotics and Automation, 2009. ICRA'09. IEEE International Conference on*, pages 3389–3394. IEEE, 2009.

K. Yamane, M. Revfi, and T. Asfour. Synthesizing object receiving motions of humanoid robots with human motion database. In *Robotics and Automation (ICRA), 2013 IEEE International Conference on*, pages 1629–1636. IEEE, 2013.

F. Yamaoka, T. Kanda, H. Ishiguro, and N. Hagita. Developing a model of robot behavior to identify and appropriately respond to implicit attention-shifting. In *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction*, pages 133–140. ACM, 2009.

A. Yamazaki, K. Yamazaki, Y. Kuno, M. Burdelski, M. Kawashima, and H. Kuzuoka. Precision timing in human-robot interaction: coordination of head movement and utterance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 131–140. ACM, 2008.

H. Yan, M. H. Ang Jr, and A. N. Poo. A survey on perception methods for human–robot interaction in social robots. *International Journal of Social Robotics*, 6(1):85–119, 2014.

H.-D. Yang, A.-Y. Park, S.-W. Lee, et al. Gesture spotting and recognition for human–robot interaction. *Robotics, IEEE Transactions on*, 23(2):256–270, 2007.

W. Yuan, E. Brunskill, T. Kollar, and N. Roy. Where to go: Interpreting natural directions using global inference. In *Robotics and Automation, 2009. ICRA '09. IEEE International Conference on*, pages 3761–3767, May 2009.

Y. Zeng, Y. Li, P. Xu, and S. S. Ge. Human-robot handshaking: A hybrid deliberate/reactive model. In *Social Robotics*, pages 258–267. Springer, 2012.

J. Zhang and A. J. C. Sharkey. Listening to sad music while seeing a happy robot face. In *Social Robotics*, pages 173–182. Springer, 2011.