

# Robots as an Interactive Media

## ABSTRACT

This paper explores robots as a newly emerging interactive media. We highlight the distinct and intriguing design challenges and interaction affordances that the physical embodiment of robots brings to the myriad of growing applications for robots in environments where there is a high level of interaction with humans, such as the home or the classroom. This motivates the need to study and evaluate human-robot interaction, much as the HCI community has done for interactions between humans and traditional computers. Towards this goal, we offer our experiences in designing robots that interact with and learn from people.

## Keywords

Human-robot interaction, sociable robots, multimodal i/o, social interaction

## INTRODUCTION

Sociable robots, autonomous robots that are specifically designed to interact with people, are an intriguing and newly emerging technology for domestic, entertainment, education, and health applications. Traditionally, autonomous robots have been targeted for applications requiring very little, if any, interaction with humans, such as sweeping minefields, inspecting oil wells, or exploring other planets. Other applications such as delivering hospital meals, mowing lawns, or vacuuming floors bring autonomous robots into environments shared with people, but human-robot interaction in these tasks is still minimal.

However, recent commercial applications are emerging where the ability to interact with people in a compelling and enjoyable manner is an important part of a robot's functionality. A new generation of robotic toys have emerged (such as Tiger Electronic's hamster-like Furby or Sony's robotic dog, Aibo) whose behavior changes the more children play with it. Although the ability of these products to interact with people is limited, they are motivating the development of increasingly life-like and socially sophisticated robots. Someday, these robotic toys might also serve an educational function for children or provide a richness of interaction that rivals that of a

beloved pet. Projects such as Aurora are exploring the use of robots to play a therapeutic role in helping children with autism [10]. Location-based entertainment applications such as theme parks or museum tour guides offer not only entertainment value but could also provide visitors with information of interest.

Corporate and university research labs are exploring applications areas for robots that assist people in a number of ways. Some companies are pursuing domestic uses. For instance, NEC is developing a small, mobile household robot that can help people interact with electronic devices around the house (e.g., TV, computer, answering service, etc.). Health-related applications are also being explored (particularly in Japan), such as the use of robots as nursemaids to help the elderly [9]. The commercial success of these robots hinges not only on their utility but also on their ability to be responsive to and interact with people in a natural and intuitive manner. Other applications include "wearable" robots such as robotic exoskeletons to help enhance the physical abilities of the elderly or disabled.

As these kinds of robots become increasingly ubiquitous in society, they must be easy for the average person to use and interact with. This raises the important question of how these sophisticated technologies should properly interact with untrained humans in a manner that is intuitive, efficient, and enjoyable to use. In the field of human computer interaction (HCI), Reeves & Nass [17] have shown that humans (whether computer experts, lay people, or computer critics) generally treat computers as they might treat other people, provided that the technology behaves in a socially competent manner. From numerous studies, Reeves & Nass argue that a social interface may be a truly universal interface given that humans have evolved to be experts in social interaction.

From these findings, we take the working assumption that attempts to foster human-robot relationships will be accepted by a majority of people *if* the robot displays rich social behavior. Similarity of morphology and sensing modalities makes humanoid robots one form of technology particularly well suited to this. If the findings of Reeves and Nass hold true for sociable robots, then those that participate in rich human-style social exchange with their users offer a number of advantages. First, people would find working with them more enjoyable and would thus feel more competent. Second, communicating with them

*LEAVE BLANK THE LAST 2.5 cm (1") OF THE LEFT  
COLUMN ON THE FIRST PAGE FOR THE  
COPYRIGHT NOTICE.*

would not require any additional training because humans are already experts in social interaction. Third, if the robot could engage in various forms of social learning (imitation, emulation, tutelage, etc.), it would be easier for the user to teach new tasks. Ideally, the user could teach the robot just as one would teach another person.

### **HCI APPLIED TO SOCIABLE ROBOTS**

While robotics researchers tackle the technical issues of building autonomous robots for these new human-centered applications, these efforts could benefit from the techniques and methodologies of the HCI community in evaluating human-robot interaction. Various task domains need to be explored including functional scenarios where robots might help a person perform a physical task, educational scenarios where a robot might help in adult training or participate in educational games for children, health-related scenarios where a robot might provide assistance to the elderly or disabled, or entertainment scenarios where the goal is a rewarding and compelling interaction.

Human-computer interaction studies as applied to human-robot interaction could be used to advance a scientific understanding of how people interact with this type of interactive technology. This, in turn, would inform the engineering of robots that interact more effectively with people. Design issues include the robot's morphology (e.g., should it be more anthropomorphic, creature-like or vehicle-like?), aesthetic appearance (e.g., should it appear organic or mechanical?), physical skillfulness, perceptual capabilities, communicative expressiveness, and its intelligence (e.g., social, emotional, and cognitive). Such design issues would be well served by human-robot interaction studies that addressed the following issues:

*Relationship issues.* What should be the nature of the human-robot relationship? Should it be more like interacting with a tool/appliance, a creature/pet, or a person (e.g., collaborator/servant)? This may depend on the application or on the person's preferred mode of interaction.

*Personality issues.* How does the person's personality impact the design of the robot? How do you design a robot to be compatible with the person's personality? Should the robot have a personality? If so, of what type and how complex?

*Cultural issues.* How do cultural attitudes impact the design? For instance, science fiction has promoted a favorable view of robots in Japanese society, whereas it has contributed to a more suspicious viewpoint in American culture. How will this impact how robots are accepted and integrated into human culture? How does this impact attitudes towards what robots should do, should not do, or cannot do? Which kinds of behavior are socially acceptable and which are inappropriate?

*Quality issues.* How does one design robots that are enjoyable, useful, and rewarding for people to interact

with? What aspects make the robot more appealing and engaging? What aspects make the robot more readily accepted and incorporated by the person? Are there aspects that make the robot intimidating or annoying?

*Naturalness issues.* How are people naturally inclined to interact with this sort of technology? In what ways will people interact with or teach it as if they would another person (using natural social cues, etc.), and in what ways might this differ? There may be advantages for both.

*User expectation issues.* What are people's implicit expectations for the robot's capabilities? For instance, do they expect the robot to be able to communicate using natural language? Do they expect the robot to understand what they are feeling? How can you design the robot to shape or calibrate the person's expectations to be commensurate with the robot's capabilities? This can mitigate the person's disappointment or frustration when interacting with the robot.

*Comparative media issues.* How does interacting with robotic technologies differ from other interactive media (such as software agents)? In what ways is it similar? Are there special affordances that a robotic media offers that could be leveraged to improve human-robot interaction? How might this compare to mixed-media applications such as merging robotics with graphical animation?

### **A DIFFERENT TYPE OF MEDIA**

Sociable robots are situated in the physical and social world of people. As opposed to a software application on a computer that a person may use only once in a while, a robot is part of the physical environment and it is likely that a person would encounter the robot on a daily basis as it goes about performing its chores. This opportunity for frequent interaction (potentially on a daily basis) over an extended period of time (potentially for years) poses some significant design challenges. Some of these challenges have to do with interacting with people, others have to do with interacting within a very complex environment. Furthermore, the physical nature of robots gives them certain interaction affordances that are distinct from those of other interactive technologies (e.g., animated or text based software agents). These design challenges need to be addressed and these affordances need to be understood for sociable robots to become a useful and rewarding part of people's daily lives.

### **Autonomy in the Real World**

For a robot, dealing with the complexity of the real world is part and parcel of performing tasks. Unlike many software agent applications, it is very difficult to restrict the robot's domain of expertise to the specific task application, given that the robot must deal with the real world when it is performing a task and when it is not (e.g., when "off-duty" the robot must tend to self-maintenance functions like recharging itself). Human society is a particularly challenging environment given its richness, its dynamic

nature, its unpredictability, and its uncertainty (imagine the complexity of everyday family life in the home to a robot). It is an environment that is not easily simplified without imposing significant restrictions (which might be unacceptable to the people that share that environment). Nonetheless, robots must perform tasks and make decisions given imperfect and partial knowledge and information. Hence, much of robotic design addresses issues of robustness, adaptability, and dealing with uncertainty – all in addition to the specific knowledge and skills required to perform a certain task.

Robots must operate within a very complex environment, yet they are quite limited in their own perceptual abilities, their motor abilities, and their intelligence (as compared to people and animals). Unlike many software agent applications with relatively “clean” interfaces for input (keyboard, mouse, etc) and output (monitor, speakers, etc), autonomous robots must cope with perceptual challenges (sensors are noisy, can drift over time, or become uncalibrated). They must cope with physical challenges (limited degrees of freedom, power consumption, torque/mass tradeoffs) and the fact that physical elements (wiring, connectors, mechanical components) fatigue, slip, or break. These limitations are exacerbated when robots roam around from location to location bumping into things, or the environmental characteristics change (lighting, wall orientation, floor texture, etc) simply because the robot wanders into a different room. Robots must somehow manage this environmental complexity to a level within their abilities. In addition, to ultimately perform tasks in the real world, the robot must map its knowledge and skills into its underlying sensor and motor modalities. For symbolically represented information (as is often the case with software agents) robotics designers now try to have the robot learn this correlation rather than by programming it in by hand.

### **Interacting with People**

When interacting with a human, sociable robots bring an interesting set of affordances. Certainly, some of these affordances are shared with other interactive media, such as embodied conversational agents [7, 13]. For instance, both agents and robots can perceive the naturally offered social cues of a human using cameras or microphones. These might include perceiving the person's tone of voice, articulated speech, facial expression, articulated gesture, body posture, and so forth. Furthermore, both have bodies (either animated or mechanical) to deliver these same social cues to a person.

To different degrees, both can share the same reference frame with a human. This is useful for exchanging deictic gestures or for establishing a shared referent through gaze direction and/or head pose. However, this is clearly more limited for a character restricted to a screen with statically mounted sensors, than for a robot whose sensors can move with it. Similarly, it is more difficult for an animated

character to establish and maintain compelling eye contact given the limitations of a planar screen. Humans are exquisitely sensitive to gaze direction and eye contact, and we have found that this ability has powerful impact on a person's sense of being engaged on a personal and direct level (although this needs to be further quantified).

There are other affordances that are particular to having a physical embodiment. For instance, robots have the ability to manipulate real objects to perform physical tasks. They are also able to locomote and move in the same physical space as people. There is also the possibility for direct physical contact between robots and people. For instance, a human might not only carry, manipulate, or wear robotic technology, but also touch it or physically interact with it as one might a pet. This introduces interesting benefits as well as possible risks. A technology is not so easily dismissed when it has the ability to proactively seek you out and come into immediate contact with you (e.g., as pets do).

### **Learning in the Human Environment**

Beyond communication and interaction, any robot that co-exists with people as part of their daily lives must be able to learn and adapt to new experiences. Ideally, people will be able to teach the robot how to do new tasks or the particulars of how to do a given task. For instance, even a task as specific as taking out the trash has a number of distinct variables, such as locating a particular trash can in a specific home, opening that style of trash can, navigating through that home and yard, and scheduling when to remove the trash.

Hence, one key challenge is to design robots that are as easy to teach as another person. Fortunately, there are many advantages that social cues and skills could offer robots that learn from people. A socially competent robot could take advantage of the same sorts of social learning and teaching scenarios that humans readily use. Below are four key challenges of robot learning, and how social, emotional, and expressive factors can be used to address them in interesting ways.

*Knowing What Matters.* Faced with an incoming stream of sensory data, a robot (the learner) must figure out which of its myriad of perceptions are relevant to learning the task. As the perceptual abilities of a robot increase, the search space becomes enormous. If the robot could narrow in on those few relevant perceptions, the learning problem would become significantly more manageable. Knowing what matters when learning a task is fundamentally a problem of determining saliency. Objects can gain saliency because of their inherent properties (motion, color, etc). The current motivational state, emotional state, and knowledge of the learner can impact saliency through contextual effects. For example, when the learner is hungry, stimuli relating to food will have higher saliency than otherwise. Objects can also become salient if they are the focus of the instructor's

attention. This would allow a person to indicate what features the robot should attend to as it learns how to perform a task. Also, in the case of social instruction, the robot's gaze direction could serve as an important feedback signal for the instructor.

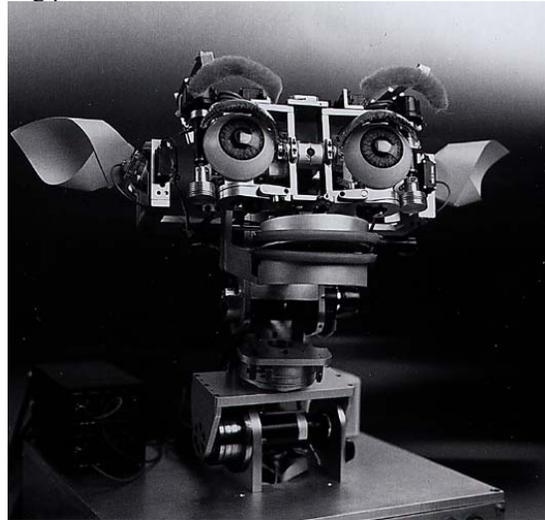
*Knowing What Action to Try.* Once the robot has identified salient aspects of the scene, how does it determine what actions it should take? As robots become more complex, their repertoire of possible actions increases. This also contributes to a large search space. If the robot had a way of focusing on those potentially successful actions, the learning problem would be simplified. In this case, a human instructor, sharing a similar morphology with the robot, could provide considerable assistance by demonstrating the appropriate actions to try. The body-mapping problem (how the robot maps its movement onto human movement) is challenging, but could provide the robot with a good first attempt.

*Correcting Errors and Recognizing Success.* Once a robot can observe an action and attempt to perform it, how can the robot determine whether or not it has been successful? Further, if the robot has been unsuccessful, how does it determine which parts of its performance were inadequate? The robot must be able to identify the desired outcome and to judge how its performance compares to that outcome. In many situations, this evaluation depends on understanding the goals and intentions of the instructor as well as the robot's own internal motivations. Additionally, the robot must be able to diagnose its errors in order to incrementally improve performance. The human instructor, however, has a good understanding of the task and knows how to evaluate the robot's success and progress. One way a human instructor could facilitate the robot's evaluation process (to recognize success and correct failures) is by providing expressive feedback. As the learner acts, the facial expressions (smiles or frowns), vocalizations, gestures (nodding or shaking of the head), and other actions of the instructor all provide feedback that allows the learner to determine whether it has achieved the goal.

*Establishing a Suitable Learning Environment.* In addition, as the instructor takes a turn, the instructor often looks to the learner's face to determine whether the learner appears confused or understands what is being demonstrated. The instructor can use the robot's expressions as feedback to control the rate of information exchange --- to either speed it up, to slow it down, or to elaborate as appropriate. By regulating the interaction, the instructor could establish an appropriate learning environment and provide better quality instruction. Finally, the structure of instructional situations is iterative: the instructor demonstrates, the student performs, and then the instructor demonstrates again, often exaggerating or focusing on aspects of the task that were not performed successfully. The ability to take turns lends significant structure to the learning episode that the learner can use to incrementally refine its performance.

## DESIGNING SOCIABLE ROBOTS

Our first robot, Kismet, is designed to be neither a tool nor an interface. One does not use Kismet to perform a task. Instead, Kismet is designed to be a robotic creature that can interact physically, affectively, and socially with humans in order to ultimately learn from them. As argued in the previous section, the ability for sociable robots to learn in a natural and intuitive way from people is a critical ability for sociable robots. Accordingly, our robot is designed to elicit interactions with the human caregiver that afford rich learning potential.



**Figure 1: Kismet, our sociable robot. Kismet has 15 degrees of freedom in its face, 3 for the eyes, and 3 for the neck. It has 4 cameras, one behind each eyeball, one between the eyes, and one in the "nose." It can express itself through facial expression, body posture, gaze direction, and vocalizations.**

Toward this goal, the design of Kismet has been strongly inspired by developmental psychology (see Figure 1). As a result, the interaction between Kismet and humans shares strong parallels in how human caregivers communicate with their infants and assist their infants' learning through similar social interactions. Somewhat like human infants, sociable robots shall be situated in a very complex social environment (that of adult humans) with limited perceptual, motor, and cognitive abilities. Human infants, however, are born with a set of perceptual and behavioral biases that serve to launch them into social interactions with their caregiver and to convey social responsiveness. Caregivers, in turn, seem to intuitively read and respond to these responses in order to adapt their behavior (e.g., slow it down, exaggerate it, and structure the interaction) to foster the infant's development. These innate abilities suggest how critically important it is for the infant to establish a social bond with his caregiver, both for survival purposes as well as to ensure normal cognitive and social development [5].

We have endowed Kismet with a substantial amount of infrastructure that we believe will enable the robot to leverage these interactions to foster its social development. These are skills and mechanisms to help it cope with a complex social environment, to tune its responses to a human, and to give the human social cues so that she is better able to tune herself to Kismet. This allows the robot to be situated in the world of humans without being overwhelmed or under-stimulated. Currently, these skills include the ability to direct the robot's attention to establish shared reference, the ability for the robot to recognize expressive feedback such as praise and prohibition, the ability to give expressive feedback to the human, the ability to take turns to structure learning episodes, and the ability to regulate interaction to establish a suitable learning environment.

Given the focus on social development and learning, it is not straightforward to directly apply standard HCI evaluation criteria to Kismet. As a result, we evaluate Kismet with respect to *instruct-ability* criteria. These are inherently subjective, yet quantifiable, measures that evaluate the quality and ease of interaction and social instruction between human and robot. They address the behavior of both partners, not just the performance of the robot. The evaluation criteria for *instruct-ability* are as follows:

- Do people intuitively read and naturally respond to Kismet's social cues?
- Can Kismet perceive and appropriately respond to these naturally offered cues?
- Does the human adapt to the robot, and the robot adapt to the human, in a way that benefits the interaction? Specifically, is the resulting interaction natural, intuitive, and enjoyable for the human; can Kismet perform well despite its

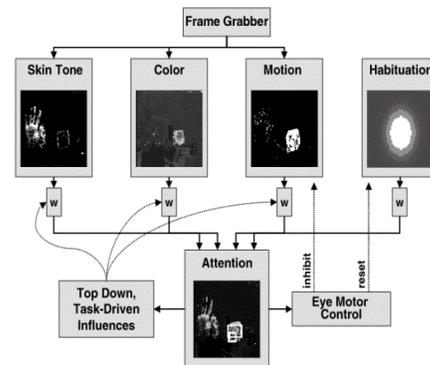
perceptual, mechanical, behavioral, and computational limitations; and is a suitable learning environment established and maintained?

- Does Kismet readily elicit structured interactions from the human that could be used to benefit learning?

During social exchanges, people send social cues to Kismet to shape its behavior. Kismet must be able to perceive and respond to these cues appropriately. By doing so, the quality of the interaction improves. Many of these social cues are offered in the context of teaching the robot. To be able to take advantage of this scaffolding, the robot must be able to correctly interpret and react to a number of social cues.

### Directing Attention

The first social cue that Kismet responds to is the ability of humans to direct Kismet's attention using natural cues [1]. This could play an important role in socially situated learning by giving the caregiver a way of showing Kismet what is important for the task and for establishing a shared reference. We have found that it is important for the robot's attention system to be tuned to the attention system of humans so that both find the same types of stimuli to be salient in similar conditions. Based on the scientific models of the visual attention system in humans as proposed by [18], we have developed a human-like set of perceptual biases in Kismet.



**Figure 2: Schematic of the robot's attention system. The robot is particularly biased to attend to saturated colors, motion, size, and skin-tone. It is inspired by J. Wolfe's theory of visual guided search in humans. Displayed images were captured during a behavioral trial session.**

The robot's attention is determined by a combination of low-level perceptual stimuli. The relative weightings of the stimuli are modulated by high-level behavior and motivational influences (see Figure 2). A sufficiently salient stimulus in any modality can preempt attention, similar to the human response to sudden motion. All else being equal, larger objects are considered more salient than

smaller ones. The design is intended to keep the robot responsive to unexpected events, while avoiding making it a slave to every whim of its environment. With this model, people intuitively provide the right cues to direct the robot's attention (shake object, move closer, wave hand, etc.).

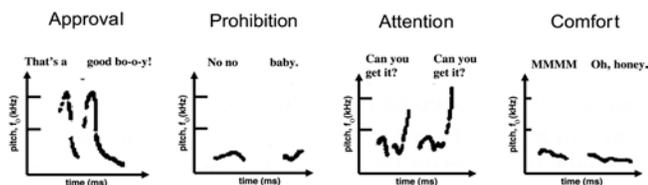
Because people and Kismet are likely to find the same visual stimuli to be attention grabbing, people can very naturally and quickly direct the robot's attention by using natural and intuitive cues (see Table 1). Kismet's attention system coupled with the active control of gaze direction provides people with a powerful and intuitive social cue for when they have succeeded in steering the robot's focus of interest.

stimulus category	stimulus	presentations	average time (s)	commonly used cues	commonly read cues
color & movement	yellow dinosaur	8	8.5	motion across center line shaking motion bringing target close to robot	eye behavior, especially tracking facial expression, especially raised eyebrows body posture, especially forward lean or withdraw
	multi-colored block	8	6.5		
	green cylinder	8	6.0		
motion only	b/w cow	8	5.0		
skin-toned & movement	pink cup	8	6.5		
	hand	8	5.0		
	face	8	3.5		
Total		56	5.8		

**Table 1: Results from a directing attention experiment. The robot quickly responds to people's attempts to direct its attention to a variety of test stimuli. The robot responds to commonly used cues such as motion and size. People intuitively read the robot's gaze direction and observe its change in visual behavior to decide when they have successfully directed the robot's attention to the desired object.**

### Recognize Affective Assessment of Human

The second social cue that Kismet responds to is tone of voice (see Figure 3). Based on the scientific study of how human infants recognize the affective intent of their caregiver's speech [12], Kismet has the ability to recognize praise, prohibition, soothing, and attentional bids from the "melody" of robot-directed speech [2].



**Figure 3: Fernald's prototypical prosodic contours for approval, attentional bid, prohibition, and soothing. Kismet is able to recognize the same affective intents.**

This serves as an important teaching cue for reinforcing and shaping the robot's behavior. Several interesting

interactions have been witnessed between Kismet and human subjects when Kismet recognizes and expressively responds to their tone of voice. They use Kismet's facial expression and body posture to determine when Kismet "understood" their intent. The video of these interactions suggests evidence of affective feedback where the subject might communicate their intent (say, an attentional bid), the robot responds expressively (perking its ears, leaning forward, and rounding its lips), and then the subject immediately responds in kind (perhaps by saying, "Oh!" or, "Ah!"). Several subjects appeared to empathize with the robot after issuing a prohibition, often reporting feeling guilty or bad for scolding the robot and making it "sad."

### Expressive Feedback

As with human infants, Kismet sends expressive feedback signals to the human caregiver, indicative of the robot's internal state. This allows the human to better predict what the robot is likely to do and to shape their responses accordingly. Kismet does this by means of expressive behavior in different modalities. It can communicate emotive state and social cues to a human through facial expressions, body posture, gaze direction, and voice [4]. We have found that the scientific basis for how emotion correlates to facial expression [15] or vocal expression [14, 6] to be very useful in mapping Kismet's emotive states to its face actuators and to its articulatory-based speech synthesizer. Further, we have found that people intuitively and naturally use Kismet's expressive feedback to tune their performance in the exchange [2, 3].

Results from various forced-choice and similarity studies suggest that Kismet's emotive facial expressions and vocal expressions are readable. In one study, seventeen subjects filled out a forced choice questionnaire to evaluate the readability of Kismet's emotive facial expressions. Most of the subjects were children 12 years of age. There were six girls, six boys, three adult men, and two adult women. There were seven pages in the questionnaire. Each page had a large color image of Kismet displaying one of seven expressions (anger, disgust, fear, happiness, sorrow, surprise, and a stern expression). The subjects could choose the best match from ten possible labels (accepting, anger, bored, disgust, fear, joy, interest, sorrow, stern, surprise). With respect to their best-choice answer, they were asked to specify on a ten-point scale how confident they were of their answer, and how intense they found the expression. The results of this study are shown in Table 2. The subjects' responses were significantly above random choice (10 percent), ranging from 47 percent to 83 percent. The misclassified images usually shared either a similar arousal (high versus low) or valence (positive versus negative) with the "correct" expression. This suggests that these affective and arousal qualities were effectively conveyed to human subjects.

	accepting	anger	interest	disgust	fear	joy	interest	sorrow	shame	surprise	% correct
anger	5.9	76.5	0	0	5.9	11.7	0	0	0	0	76.5
disgust	0	17.6	0	70.6	5.9	0	0	0	5.9	0	70.6
fear	5.9	5.9	0	0	47.1	17.6	5.9	0	0	17.6	47.1
joy	11.7	0	5.9	0	0	82.4	0	0	0	0	82.4
sorrow	0	5.9	0	0	11.7	0	0	83.4	0	0	83.4
shame	7.7	15.4	0	7.7	0	0	0	15.4	53.8	0	53.8
surprise	0	0	0	0	0	17.6	0	0	0	82.4	82.4

Forced-Choice Percentage (random=10%)

**Table 2: Results from forced choice experiments to assess the readability of Kismet's facial expressions.**

### Regulating Interactions

In addition to emotive expressions, Kismet employs communicative facial displays (such as envelope displays—those that regulate the exchange of turns in a dialog) to regulate the exchange of speaking turns between human and robot. These include both eye movements as well as postural and facial displays. We found that people intuitively and naturally use Kismet's expressive feedback to entrain their performance to the robot during vocal turn-taking exchanges [8, 3]. We also have found that both the human and robot benefit from this: the person enjoys the easy interaction while the robot is able to perform effectively within its perceptual (i.e., auditory), computational, and behavioral limits.

To investigate Kismet's performance in engaging people in proto-dialogues, we invited three naive subjects to interact with Kismet. They ranged in age from 25 to 28 years of age. One male and two females participated in the experiment. All were professionals. They were asked simply to talk to the robot as they might engage a pre-linguistic infant (Kismet babbles, but does not employ natural language yet). Their interactions were video recorded for further analysis.

Often the subjects begin the session by speaking longer phrases and only using the robot's vocal behavior to gauge their speaking turn. They also expect the robot to respond immediately after they finish talking. Within the first couple of exchanges, they may notice that the robot interrupts them, and they begin to adapt to Kismet's rate. They start to use shorter phrases, wait longer for the robot to respond, and more carefully watch the robot's turn-taking cues. The robot prompts the other for her turn by craning its neck forward, raising its brows, and looking at the person's face when it's ready for her to speak. It will hold this posture for a few seconds until the person responds. Often, within a second of this display, the subject does so. The robot then leans back to a neutral posture, assumes a neutral expression, and tends to shift its gaze away from the person. This cue indicates that the robot is about to speak. The robot typically issues one utterance, but it may issue several. Nonetheless, as the exchange proceeds, the subjects tend to wait until prompted.

		time stamp (min:sec)	time between disturbances (sec)
subject 1	start @ 15:20	15:20 – 15:33	13
		15:37 – 15:54	21
		15:56 – 16:15	19
		16:20 – 17:25	70
	end @ 18:07	17:30 – 18:07	37+
subject 2	start @ 6:43	6:43 – 6:50	7
		6:54 – 7:15	21
		7:18 – 8:02	44
	end @ 8:43	8:06 – 8:43	37+
subject 3	start @ 4:52 min	4:52 – 4:58	10
		5:08 – 5:23	15
		5:30 – 5:54	24
		6:00 – 6:53	53
		6:58 – 7:16	18
		7:18 – 8:16	58
		8:25 – 9:10	45
	end @ 10:40 min	9:20 – 10:40	80+

**Table 3: Data for the vocal turn-taking experiment.**

Before the subjects adapt their behavior to the robot's capabilities, the robot is more likely to interrupt them. There tend to be more frequent delays in the flow of "conversation," where the human prompts the robot again for a response. Often these "hiccups" in the flow appear in short clusters of mutual interruptions and pauses (often over two to four speaking turns) before the turns become coordinated and the flow smoothes out. By analyzing the video of these human-robot "conversations," there is evidence that people entrain to the robot (see Table 3). These "hiccups" become less frequent. The human and robot are able to carry on longer sequences of clean turn transitions. At this point the rate of vocal exchange is well matched to the robot's perceptual limitations. The vocal exchange is reasonably fluid. Table 4 shows that the robot is engaged in a smooth proto-dialogue with the human partner the majority of the time (about 82 percent).

	subject 1		subject 2		subject 3		average
	data	percentage	data	percentage	data	percentage	
clean turns	35	83%	45	85%	83	78%	82%
interrupts	4	10%	4	7.5%	16	15%	11%
prompts	3	7%	4	7.5%	7	7%	7%
significant flow disturbances	3	7%	3	5.7%	7	7%	6.5%
total speaking turns	42		53		106		

**Figure 4: Evidence of entrainment in human-robot vocal turn taking.**

### SUMMARY

In this paper we have explored robots as a newly emerging interactive media. To support this view, we have highlighted the distinct design challenges and interactive affordances that are characteristic of this physically animated media. We have advocated applying similar techniques and methodologies developed in the HCI community to characterize and understand human-robot interaction. There will be some similarities in how people

interact with software agents and with robots, but the physical embodiment of robots will bring differences as well. We have highlighted a few of our research efforts to illustrate how these design issues and affordances have been addressed on our sociable robot, Kismet.

#### ACKNOWLEDGMENTS

The author gratefully acknowledges the creativity and ingenuity of the members of the Robotic Presence Group at the MIT Media Lab and of the Humanoid Robotics Group at the MIT Artificial Intelligence Lab. Work relating to Kismet was funded by NTT and DARPA contract DABT 63-99-1-0012.

#### REFERENCES

1. Breazeal, C. & Scassellati, B., "A Context-Dependent Attention System for a Social Robot," *Proceedings of the 16<sup>th</sup> International Conference on Artificial Intelligence* (Stockholm, Sweden 1999). 1146—1151.
2. Breazeal, C. & Aryananda, L., "Recognition of Affective Communicative Intent in Robot-Directed Speech," *Proceedings of the 1<sup>st</sup> IEEE-RAS International Conference on Humanoid Robots* (Cambridge MA 2000).
3. Breazeal, C., "Regulation and Entrainment in Human-Robot Interaction," *Proceedings of the 7<sup>th</sup> International Symposium on Experimental Robotics* (Honolulu, HI 2001).
4. Breazeal, C., "Believability and Readability of Robot Faces," *Proceedings of the 8<sup>th</sup> International Symposium on Intelligent Robot Systems* (Reading UK 2000), 247—256.
5. Bullowa, M. (ed.), *Before Speech: The Beginning of Interpersonal Communication*. Cambridge University Press, Cambridge UK. 1979.
6. Cahn, J., *Generating Expression in Synthesized Speech*. S. M. Thesis, Massachusetts Institute of Technology Department of Media Arts and Sciences, Cambridge MA, 1990.
7. Cassell, J., Sullivan, J., Prevost, S. & Churchill, E. (eds.), *Embodied Conversational Agents*, MIT Press, Cambridge MA. 2000.
8. Cassell, J., "Nudge nudge wink wink: Elements of face-to-face conversation for embodied conversational agents" *Embodied Conversational Agents*, Cassell, J., Sullivan, J., Prevost, S. & Churchill, E. (eds.) MIT Press, Cambridge MA, 1—27, 2000.
9. Dario, P. & Susani, G., "Physical and psychological interactions between humans and robots in the home environment," *Proceedings of the First International Symposium on Humanoid Robots* (Tokyo Japan 1996), 5—16.
10. Dautenhahn, K. "Robots as social actors: Aurora and the case of autism." *Proceedings of the 3<sup>rd</sup> International Cognitive Technology Conference*, (San Francisco, CA 1999), 359-374.
11. Dautenhahn, K. "The Art of Designing Socially Intelligent Agents: Science, Fiction, and the Human in the Loop." *Applied Artificial Intelligence Journal* 12, 7—8 (1998), 573-617.
12. Fernald, A., "Intonation and Communicative Intent in Mother's Speech to Infants: Is the Melody the Message?," *Child Development*, 60, 1497—1510, 1989.
13. Lester J., Towns, S., Callaway, S., Voerman, J. & Fitzgerald, P., "Deictic and emotive communication in animated pedagogical agents," *Embodied Conversational Agents*, Cassell, J., Sullivan, J., Prevost, S. & Churchill, E. (eds.) MIT Press, Cambridge MA, 123—154, 2000.
14. Murray, I. & Arnott, L., "Toward the Simulation of Emotion in Synthetic Speech: A Review of the Literature on Human Vocal Emotion," *Journal Acoustical Society of America*, 93(2), 1097—1108, 1993.
15. Smith, C. & Scott, H., "A Componential Approach to the Meaning of Facial Expressions," *The Psychology of Facial Expression*, J. Russell & J.M. Fernandez-Dols (eds.), Cambridge University Press, Cambridge UK, 229—254, 1997.
16. Mackay, W.E. Ethics, lies and videotape, in *Proceedings of CHI '95* (Denver CO, May 1995), ACM Press, 138-145.
17. Reeves, B. & Nass, C., *The Media Equation*. CLSI Publications, Stanford CA, 1996.
18. Wolfe, J., "Guided Search 2.0: A revised model of visual search," *Psychonomic Bulletin and Review*, 1(2), 202—238, 1994.