6

Nonverbal Interaction

What is covered in this chapter:

- The role of nonverbal communication in interactions between people—how communication is enhanced by facial expressions, hand gestures, body posture, and sounds.
- The importance of interpreting, using, and responding to nonverbal cues in the appropriate way, both to successful human–robot interactions and to generate a positive perception of robots.
- Nonverbal communication channels that are unique to robots, as well as channels that replicate those commonly used by humans.
- How robotic sounds, lights, and colors or physical gestures with arms, legs, tails, ears, and other body parts can be effective for communicating with people.

When we think of what it means to communicate with someone face to face, the first thing that comes to mind is often the content of our speech—what we are saying to each other—rather than the manner in which such content is delivered. Just for a minute, though, imagine speaking face to face with someone without the ability to look at the person or use gestures. Not only would you be uncomfortable, but you might also have difficulty getting the intended meaning across. Moreover, without the nonverbal "channel," it seems harder to establish a strong connection with the person, particularly when you communicate with a stranger.

This chapter covers this unspoken (i.e., nonverbal) component of our social interactions, both with other humans and specifically with robots. Section 6.1 highlights the different functions that nonverbal communication plays. Section 6.2 dives deeper into the different kinds of nonverbal behavior, such as gaze and mimicry. Finally, Section 6.3 explicitly focuses on how robots can read and generate nonverbal behavior.

As shown in the previous example, people constantly and seemingly automatically give off and pick up on a variety of nonverbal cues while interacting. These cues are used to interpret the nuances of meaning, emotion, and intention in others. Nonverbal cues are such an important aspect of human communication that being unable to produce and decipher them appropriately makes interaction quite challenging. Anyone may experience a sense of bewilderment when they go to another country—we may find it difficult

to summon the waiter to give us the bill or might struggle to read another person's face in order to understand what he or she is feeling. The importance of nonverbal cues is acutely experienced by people with disorders such as autism, who have difficulty noticing and interpreting nonverbal social cues in others. On the other hand, being sensitive to nonverbal cues can improve one's understanding of an interaction. For example, researchers who have used "social sensors" to measure aspects of nonverbal behavior, such as gaze and rhythmicity, can predict which people will exchange cards at a conference (Pentland and Heibeck, 2010) or which couples will break up within a six-year period, based on thin slices of nonverbal behavior (Carrere and Gottman, 1999).

Even in the earliest social robot designs, nonverbal cues that are present in human interaction have been actively used to enrich interactions with the robot. They are typically used in combination with speech to provide supplemental information on the robot's internal state or intentions. Kismet, one of the first social robots (see Figure 2.4), used postural cues, such as pulling back or leaning forward, to express affect and engage people in interaction (Breazeal, 2003). Keepon, a minimalist social robot (see Figures 2.7 and 4.8), uses gaze and reactive motion to express attention and affect (Kozima et al., 2009). Many robots are also capable of engaging in joint attention to signal engagement with the user and a shared task. Next, we discuss the functions and types of nonverbal cues and their uses in human–robot interaction (HRI).

6.1 Functions of nonverbal cues in interaction

Nonverbal cues allow people to communicate important information "between the lines." They add a further layer of information to human (and human–robot) interaction, adding to what is being communicated linguistically. Through nonverbal communication, people can signal mutual understanding, shared goals, and common ground. They can communicate thoughts, emotions, and attention. And they can do so in a more subtle, indirect manner than through verbal expression.

In psychology, nonverbal communicative cues, such as eye gaze, body posture, or facial muscle activity, are often studied as implicit indicators of affect toward a person or an object. Many of the nonverbal signals we convey are expressed automatically without much thought or even entirely unconsciously. Therefore, nonverbal cues are often believed to be unfiltered and more genuine, revealing people's "true" attitudes. For instance, your body language can communicate a message very different from your speech. Think of an acquaintance you do not like very much. Although you might greet this person in a friendly manner and start a seemingly amicable chat, your nonverbal cues might give away your true feelings. You might look at the person more briefly, frown rather than smile, and avoid physical contact while not even being aware that your nonverbal cues are incongruent with your verbal chitchat.

Nonverbal cues are equally important for HRI. Nonverbal cues produced by people when interacting with a robot can indicate whether a person is enjoying the interaction and whether the person likes the robot or not. They can therefore act as a measure or cue of attitude or engagement and be used to guide the robot's behavior. Even in the HRI context, verbal and nonverbal cues might be contradictory. For example, people may verbally express positive ideas about a robot while the nonverbal cues suggest they are tense or anxious while interacting with it. HRI may also be affected by the way robots produce nonverbal cues. For example, an interaction can appear awkward when the robot produces gestures that do not match the rhythm or meaning of its speech or when it does not respond appropriately to people's nonverbal cues. Early research on HRI focused mainly on speech as the most obvious mode of communication for robots, but researchers now agree that nonverbal cues are central to HRI, and their implementation is widely accepted as a prerequisite for smooth and successful interaction between humans and robots. To illustrate, think of human eye gaze during a conversation. Eye gaze occurs automatically, without much thought, but at the same time, it signals shared attention—that both people are talking about the same thing—and acknowledges the conversation partner. When speaking to a robot, we would expect the robot to turn its head toward us and make eye contact with us, showing that it is attending to what we say. A robot that displays such nonverbal behavior will make the interaction seem more natural and smooth. Conversely, we notice immediately when some of this "social glue" is absent—we can sense that something is going wrong, even though it might be difficult to pinpoint exactly what is missing. When the robot stares straight ahead and does not acknowledge our presence or spoken requests, the interaction breaks down.

As with all information, nonverbal communication always occurs in a specific context, which renders the respective nonverbal signals appropriate or not. This context may be restricted by specific social and cultural norms. For example, in Western societies, people shake hands to greet each other formally, whereas a respectful greeting in Japan is performed by bowing. Even the degree to which one person bows to another signals social status and hierarchy. This might be almost imperceptible to the naive observer, but it is immediately obvious to those who understand the relevant cultural norms. Similarly, a conversation with a person from a Western society would naturally include continuous eye contact or even physical touch. However, this might be interpreted as threatening or rude in another cultural context. Such social and cultural differences are being taken up in recent HRI research on designing culturally sensitive interactions, such as investigating the importance of nonverbal cues for the cross-cultural deployment of social robots. For example, researchers from the United Kingdom and Japan have worked together to develop culturally competent care robots, a task that involves developing cultural knowledge representations, culturally sensitive planning and execution, and culturally appropriate multimodal HRI (Bruno et al., 2017). Designing HRI that meets social norms and cultural expectations might mean

the difference between a successful product and a wasted investment (see Figure 6.1).

Figure 6.1
Culturally appropriate nonverbal cues can make communication between people and robots more natural and pleasant.

6.2 Types of nonverbal interaction

Although we exhibit and experience nonverbal cues in several modalities at once—such as sound, movement, and gaze—it might be worthwhile to consider each channel of communication separately when trying to implement nonverbal signals into HRI. Understanding the functions and effects of various nonverbal cues allows us to then combine them as needed for different tasks and effects in HRI.

### 6.2.1 Gaze and eye movement

Imagine you are conducting a job interview, and the job candidate responds to your inquiries without looking at you, staring only at the desk in front of them. Even while you are sketching a graph on the whiteboard, the job candidate does not follow your gaze toward what you are drawing. Would you hire the person? Probably not, because this type of gaze behavior would likely come across as a lack of interest in you and what you are talking about.

Gaze is a subtle and important cue for managing social interaction. Gaze signals interest, understanding, attention, and people's ability and willingness to follow the conversation. Beyond their social function, gaze and eye movements also facilitate functional interactions and collaboration, such as handing an object to someone or calling someone's attention to the next tool needed in a task. Eye-tracking technology can be used to assess gaze patterns and provide insights into information processing and human cognition. Pragmatically, analyzing gaze patterns can also help to ensure that a given task has been completed smoothly. Gaze can also be a way of soliciting and keeping another person's attention during an interaction. For instance, gaze can be a way to manage turn-taking in interactions; by looking from one person to another, the speaker might suggest whose turn it is to speak next.

A particularly well-established component of gaze behavior in human interaction is joint attention. *Joint attention* refers to interaction partners attending to the same area or object at the same time. The significance of this behavior for human development starts in early childhood, when joint attention is a major scaffold for learning. The ability to attend to the same object at the same time with an adult caregiver is an important prerequisite for infants' ability to learn new words and behaviors (Yu and Smith, 2013), whereas the inability to perform joint attention can lead to developmental difficulties (Charman et al., 2000). Joint attention in adult communication can also signify interest and deep involvement in the interaction and is important for collaborative tasks where actors need to coordinate their activities. To achieve joint attention, the timing and synchrony of gaze behavior are important aspects to consider.

The eyes are a window to the soul, or in this case, they unconsciously reveal how much you like your interaction partner. Pupil dilation is controlled by the autonomic nervous system, as are uncontrollable reactions such as an increase in heart rate or goose bumps. When people see physically attractive others, their pupils automatically dilate. This also works the other way: people judge faces with larger pupils as more attractive than those with more visible irises. Pupil dilation can be used on robots to give the impression that the robot is attracted to the user (see Figure 6.2).

Figure 6.2 Pupils signal attraction, even in robots.
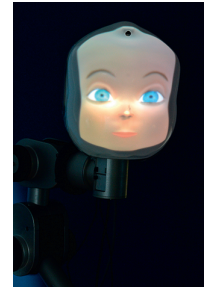
Joint attention has been incorporated into HRI in several ways: Imai et al. (2003) used it as a way of scaffolding smoother communication with people so that they know what the robot is talking about, both in conjunction with and without speech. Joint attention has also been studied as a fundamental capability of robots that want to learn from humans, particularly humanoid robots (Scassellati, 1999). Finally, joint attention with robots has been studied in interactions with children who have autism, with the aim of using the robot to assist them in developing this important social skill.

When used in HRI, robot gaze cues most often produce similar effects as they would in human interactions. This may be because researchers have used human gaze behavior to derive models of gaze behavior for robots, and they have shown that the resulting gaze cues can be used to lead people to take on different conversational roles as addressees, bystanders, or nonparticipants (Mutlu et al., 2012). In a multiparty interaction, a robot can use its gaze to control who will be the next person to talk (Mutlu et al., 2009). Andrist et al. (2014) used face-tracking movements to engage in mutual gaze and purposeful gaze aversions in an HRI study to show that such cues can make a robot seem more intentional and thoughtful. Mutlu et al. (2006) also showed that a robot's gaze cues, modeled on those of humans, used in the course of telling a story affected how well people remembered the story's content; the people with whom the robot kept gaze contact could recall more details from its story. Robot gaze can therefore be a powerful way to manage interactions with one or more people (see Figure 6.3).

6.2.2 Gesture

Following speech, gesturing is perhaps the most apparent way of providing information during an interaction. Gestures can function in place of or along with speech and are often categorized based on their role in communication. Deictic gestures involve pointing to specific things in the environment and can be important for establishing joint attention. Iconic gestures often go along with speech, further supporting and illustrating what is being said. For example, opening your arms wide while saying you are holding a big ball would be an iconic gesture, as would smoothly moving your hand upward

Figure 6.3 The eyes of robots are often designed to pitch and yaw, allowing a robot to use gaze as an effective communication channel. Here, iCub (2004–present) gives a good impression of attending to the ball in its left hand.
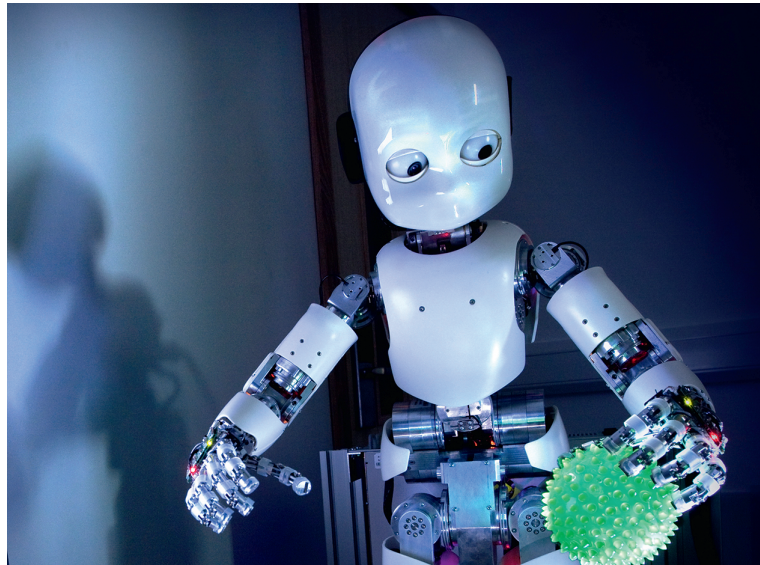

Figure 6.4 A Pepper (2014–present) robot using hand gestures to accompany its speech. Without these automatically generated beat gestures, the robot's speech would appear less natural. (Source: Pepper robot by SoftBank Robotics)

while explaining how your airplane took off. Symbolic gestures, such as waving hello or goodbye, can carry their own meaning, with or without accompanying speech. Finally, beat gestures are used to go along with the rhythm of speech and look like moving one's arms while speaking as if conducting an invisible orchestra (see Figure 6.4). Gestures can also be used for emphasizing particular moments during speech, such as lifting your hands up while saying "what?" when you are surprised by something.

Gestures are likewise a powerful way of enhancing spoken communication in HRI. A robot may be designed to gesture through its arms and hands or other body parts, such as its head, ears, or tail. The shape, timing, naturalness, and smoothness of gestures can also affect people's perceptions and understanding (Bremner et al., 2009). Salem et al. (2013) showed that including gestures along with speech in HRI led to the ASIMO humanoid robot being perceived as more anthropomorphic and likable, with participants expressing greater willingness to interact with the robot later on than when the robot communicated through speech alone. Interestingly, this study also showed that the use of gestures performed incongruently with speech led to even more pronounced

positive effects in evaluations of the robot, although it had a negative effect on task performance. Gestures should therefore be used carefully in the design of robots, and their effects should be tested in studies with humans to gauge their effects on specific interactions.

The generation of co-speech gestures for artificial agents, such as online characters or robots, is now largely done through machine learning. Gesture-generation software is trained by feeding it hundreds of hours of video recordings of people talking, allowing machine learning to build a model that generates gestures that closely fit spoken communication. For now, these models will not always generate a gesture that fits the meaning of what is being said, but the dynamics of gestures themselves are virtually indistinguishable from human co-speech gestures (Yoon et al., 2022).

6.2.3 Mimicry and imitation

Another aspect of nonverbal interaction that has been given much attention in the human-interaction literature is mimicry and imitation. By *mimicry*, we mean the unconscious replication of the behavior of another person, and by *imitation*, we mean the conscious replication of another's behavior (Genschow et al., 2017). Mimicry and imitation are performed not only by humans but also by primates (hence the notion of "aping someone") and are considered basic social capabilities. Mimicry is so central to human cognition that it has been found to have a neurological basis. The *mirror neuron system* in the primate brain contains clusters of mirror neurons (Rizzolatti and Craighero, 2004). These neurons fire both when observing someone doing an action you also know how to do, such as picking up a grape, or when you are doing that action yourself, and they are believed to be responsible for facial mimicry, the unconscious and automatic mimicking of the facial expressions of others Rymarczyk et al. (2018).

> Researchers in Japan found a band of macaques that all wash their sweet potatoes in a stream. This behavior was traced to a female member of the troop, who may have initially done this once by accident, and then others copied her when they realized that washing the potatoes produced a less gritty and more pleasing meal, so they continued the practice. Observations of this kind have led to the claim that animals, not only humans, have "culture" (Whiten et al., 1999; De Waal, 2001).

In humans, mimicry and imitation have multiple developmental functions. In early childhood development, mimicry and imitation provide a common way to learn new behaviors and culturally relevant social norms. Children use mimicry to learn to do things in particular ways—such as talking with a British accent or making expressions similar to those of a family member. As adults, we can also use imitation to blend into our social and cultural surroundings, such as gesturing more emphatically when we are speaking Italian or visiting

Italy. As such, imitation and mimicry can be important ways of developing signs of in-group identity.

Mimicry, as a largely automatic behavioral response, also has many significant social functions; one is that it indirectly signals positive affect and liking for an interaction partner. If two people use the same gestures or adopt the same postures during a conversation, it is usually because they have established a positive relation in that interaction. Similarly, when people's nonverbal cues are out of sync and not mirroring each other, you can sense that the communication is not running smoothly. Mimicry, as a subtle nonverbal cue, can thus be a helpful signal to interpret, for instance, in the context of dating or job interviews.

> Mimicry's significance in establishing a social relationship with another person makes it possible for its manipulation to function as a tool for persuasion. In studies of the "chameleon effect," Chartrand and Bargh (1999) found that subtle mimicry of a person's gestures and posture can help that person persuade an interaction partner to agree with his or her suggestions. For example, if you sit with your right leg crossed over your left, and your interaction partner subtly adopts that position, too, before telling you that Candy A tastes better than Candy B, you are more likely to choose to try Candy A over Candy B than if the person had not mimicked your posture (see Figure 6.5). However, this effect is time dependent. If you notice your conversation partner mimicking you, either because they are too obvious about it or too late in their timing, their intentions will backfire because you may see them as manipulative or insincere.

Various aspects of imitation and mimicry have been implemented and evaluated in the design of robots. There is a large and growing collection

Figure 6.5 Similar to a chameleon adjusting its color to the environment, the *chameleon effect* refers to mimicking a person's gestures or facial expressions.

This material has been published by Cambridge University Press as Human-Robot Interaction by
Christoph Bartneck, Tony Belpaeime, Friederike Eyssel, Takayuki Kanda, Merel Keijsers, and Selma Sabanovic.
ISBN: 9781009424233 (https://www.cambridge.org/9781009424233).
This pre-publication version is free to view and download for personal use only. Not for re-distribution, re-sale or use in derivative works.

of literature on robot learning by imitation, in which robots in some way record and then reproduce actions performed by humans (Argall et al., 2009; Mostafaoui et al., 2022). Riek et al. (2010) developed an apelike robot that mimicked users' head gestures, and their findings suggest this made a positive contribution to people's interactions with the robot, although these gestures were not always clear to participants.

If we combine what we know about mimicry (see Section 6.2.3) and posture (which will be discussed shortly, in Section 6.2.5) from human psychology, we can design robots that are able to display certain types of behaviors (e.g., leaning in) to affect how people behave and, therefore, how they feel. For example, Wills et al. (2016) showed that a robot that mimicked people's facial expressions and displayed socially contingent head poses received more monetary donations than a robot that did not display such behavior. Imitation and mimicry can therefore be used as both conscious and unconscious social cues in HRI to improve interaction and persuade people to follow the robot's suggestions.

6.2.4 Touch

Touch is a nonverbal cue that is often involved in close interactions among people, such as those between friends or between caregivers and patients. We often use touch deliberately, for example, to calm down someone who is agitated or to console someone who is sad. We also often incidentally touch people we feel attracted to or whom we like. It turns out that these people often also like us more when this happens. Both deliberate and incidental touch can therefore have beneficial effects, particularly when the interaction partners are part of the same social group. It is important, however, to know when and how it is appropriate to touch someone.

In everyday life, touch is sometimes used deliberately to achieve a goal. According to the so-called Midas effect, waiters and waitresses get a higher tip if they happen to incidentally touch their customers before they pay for their meal (Crusco and Wetzel, 1984). Touch does not always have positive effects, however, particularly when people who identify with different social groups are interacting with each other. In this case, touch may even lead to more negative feelings about the interaction partner. Incidental touch has also been shown to lead to a reduction in more indirect, but not direct, forms of prejudice against an out-group (Seger et al., 2014). Results on the effects of touch between human groups are therefore mixed, and it is interesting to consider what role touch might play in interactions between humans and robots, which may represent a new social group in the society of the future.

Physical interaction with robots is relatively uncommon for a number of reasons. One significant issue is the lack of safety during haptic interactions, with simple actions like hugging, holding, or shaking hands already forming a safety concern. Despite the fact that many robots are explicitly designed for HRI, tactile or haptic interaction has in many cases been an afterthought. For example, Aldebaran Robotics' Nao and Pepper robots, while widely sold as

Figure 6.6 Telenoid (2010–2013) is a haptic robot that is designed to be hugged. Studies on whether this is a form of interaction people are comfortable with are ongoing. (Source: Hiroshi Ishiguro)

Figure 6.7 The Haptic Creature (2005–2013) was designed by Steve Yohanan to study the role of affective touch in social HRI. (Source: Steve Yohanan)

commercial social robots, both have pinch points under the arms and near the hips that can painfully trap fingers. Only a limited number of robots are explicitly designed to allow physical human–robot interactions (see Figure 6.6), with most robots prone to suffer breakage when being manipulated or handled with forces that would not be uncommon in interactions between people.

The few studies on touch in HRI that are available in the literature demonstrate the need for more empirical work on this nonverbal cue (Van Erp and Toet, 2013; Willemse et al., 2016). On the positive side, tactile interaction with animal-like robots, such as Paro or the Haptic Creature (Figure 6.7), show that people can feel less stressed and anxious when they initiate such interactions (Shibata, 2012; Yohanan and MacLean, 2012). Chen et al. (2014) showed that people did not mind being touched by a robot in a nursing scenario, but they evaluated functional touch (e.g., to clean their arm) more positively than affective touch (e.g., to comfort them). In contrast, a recent study by Wullenkord et al. (2016) explored the negative consequences of touch in an interaction with the robot Nao. Participants reported their attitudes toward a Nao robot, then had to touch the robot as part of a task. After the task, they reported their attitudes and social judgments about the robot again. Overall, contact improved the participants' attitudes, such that people expressed more positive and less negative attitudes after the touch interaction as compared to one without touch. However, people who had particularly negative emotions toward robots at the onset of the study experienced the opposite effect and had more negative perceptions after they touched the robot.

Touch is an integral part of natural human–robot interactions, for example, in functional tasks such as object handovers and manipulation and in social tasks such as a handshake for greeting. In both functional and social uses, we need to keep in mind the psychological implications of incidental or deliberate touch, whether it is being touched by a robot or having to touch a robot.

6.2.5 Posture and movement

People also communicate through their full-body posture and the way in which they move. Along with facial expressions, postures can be used to signal a person's emotional state. Slow movements, drooping shoulders, and lethargic gestures all suggest a downcast state of mind, whereas fast movements and an upright bearing are signs of a positive attitude. These types of postural cues are particularly important when a person's face is not visible, but they can also provide additional cues to a person's state of mind even when we can see the person's facial expression. Researchers have found that people can interpret these types of nonverbal cues not only when they see the whole body of the person but also in minimalist light dot displays that depict a person's movements (Alaerts et al., 2011).

The way we pose can signal attention, engagement, and attraction in an interaction between humans. People might be displaying a defensive posture by holding their arms in front of them, whereas open arms are a clear invitation for engagement, perhaps even a hug. How we are posed in relation to other

people can also provide valuable information; if two people are sitting with their knees toward each other, it shows willing engagement, whereas if one person is turned partly away from the other, it can show a desire to discontinue the interaction.

> The Thrifty Faucet (2009) is a simple interactive prototype that uses its posture to communicate 15 lifelike motion patterns, including seeking, curiosity, and rejection, to users. The aim is to enable communication with users about more sustainable water use (Togler et al., 2009).
>
>
>
> (Source: Jonas Togler)

Bodily postures can provide an additional layer of expressiveness to robots (see Figure 6.8). To illustrate, when a robot lacks expressive facial features, the body can be used as the primary way to communicate emotions. Beck et al. (2010) showed that affective body postures can improve people's understanding of a robot's emotional state. A robot's posture can be used to express emotion and, through that, affect the emotions of onlookers. Xu et al. (2014) showed that people were not only able to interpret the affective body postures of robots but also that they adopted the emotions they thought the robots were showing.

Robot designers have furthermore realized that micromovements, barely perceptible motions, can convey the impression that the robot is lifelike (Yamaoka et al., 2005; Ishiguro, 2007; Sakamoto et al., 2007). These micro-movements are often implemented as small, random perturbations to the robot's actuators. Such lifelike animations can also be used to signal the robot's

Figure 6.8 A Nao robot (2008–present) using body postures to express emotions, morphing between sad (left) and fearful (right). (Source: Beck et al. (2010))

internal state; for example, the velocity or amplitude of the motion signals the excitement level of the robot (Belpaeme et al., 2012). This approach has been successfully used on petlike small robots (Cooney et al., 2014; Singh and Young, 2012).

6.2.6 Interaction rhythm and timing

The temporal nature, or "timing," of communicative cues carries its own significance in interaction. In verbal communication, we refer to this as *turn-taking* among interaction partners. Nonverbal cues (e.g., gaze, gesture) can support this turn-taking by guiding attention to the appropriate interaction partner or signifying the end of a turn. Establishing synchronized temporal patterns of interaction can further scaffold the communicative and collaborative success of an interaction.

The "rhythmicity" and "synchrony" of an interaction provide a largely unconscious but crucial component of human communication. To understand what we mean by interaction rhythms, think about human interaction as a coupled system working together. In order for two people to be able to communicate and work effectively, they need to become "rhythmically entrained" to each other's actions—to be doing things not necessarily at the same time but to the same beat. Like in dance, rhythmicity enables people to be more attuned to each other's communicative cues, to be looking, speaking, and moving at the right time to enable clear and smooth communication among the two partners (Warner et al., 1987). Although often unconscious, the effects of rhythmicity on interaction are significant: being out of synchrony can imply that interaction partners have missed important social signals and are therefore unable to interpret each other's behavior; it can also lead to a more negative interaction outcome and a less positive attitude toward the other person.

Michalowski et al. (2007) showed that a robot that is rhythmically entrained with a human interaction partner is considered more lifelike than a robot that is behaving rhythmically but is not synced with the human. They also showed that people are more likely to interact for a longer time with a rhythmically entrained dancing robot. Rhythmicity in interaction can also be useful in supporting turn-taking and collaboration in teams, including anticipation of people's behaviors and when they will show those behaviors (Hoffman and Breazeal, 2007). Finally, Siu et al. (2010) showed that listening to highly rhythmic music while performing robotic surgery can improve the performance of the human–robot surgical team. These findings suggest that rhythmicity in HRI can improve both the perceived quality of the interaction and the chances of a successful outcome.

6.3 Nonverbal interaction in robots

6.3.1 Robot perception of nonverbal cues

Standard pattern-recognition techniques are used to allow robots to perceive and identify human nonverbal cues. Posture and gesture recognition are well

studied. Typical systems use cameras, depth cameras, or sensors carried by the user to record a time series of data. Although software could be written to recognize a limited number of gestures, instead it is common to use machine learning to train the system to recognize gestures and other nonverbal cues. To achieve this, a database is collected of, for example, people showing different gestures. Typically, thousands or even millions of data points are needed, and each needs to be labeled, meaning that for each data point, we need to note what it is showing. Is it a person waving, pointing, or beckoning? Next, a classifier is trained on the labeled data; this is often an iterative process in which the classifier's performance improves when more data are processed. Once the performance is sufficient for the application, the classifier is deployed on the robot (Mitra and Acharya, 2007).

These basic perception techniques are used to allow HRI researchers to evaluate whether people are actually engaged in interactions with robots. Unlike typical human interaction, where it is expected that the human partner will be attentive and engaged, in HRI, users sometimes do not attend to what the robot says and signals. Thus, perceiving the "engagement" of users is a crucial step for enabling robots to create a successful interaction. Rich et al. (2010) developed a technique to integrate the detection of cues such as eye contact and back-channeling to identify whether a user is engaged in interaction. Sanghvi et al. (2011) analyzed affective postures and body motion to detect engagement with a robotic game companion.

Although the constant advancement of technology allows for the improvement of robotic perception capabilities, researchers also add special equipment to the robot, such as eye trackers and motion-capture systems, to provide data on nonverbal cues relevant for interaction. For tactile interaction, there has been some research in the robotics field in which film-type piezoelectric polymer sensors were inserted in thin and thick silicone rubber (Taichi et al., 2006) or where flexible capacitive sensors or skin with nano hairs is added to the robot's body to allow the detection of tactile events with high sensitivity and high temporal and spatial resolution (Yogeswaran et al., 2015).

6.3.2 Generating nonverbal cues in robots

Generating gestures and other nonverbal cues is not trivial in robots. The nonverbal cues need to be contingent on the interaction: if the user snaps her fingers, the robot needs to blink immediately. Nonverbal cues also need to be coordinated with each other and with other cues, including verbal interaction, in terms of both the semantic meaning and the timing of execution. HRI poses particular challenges for the perception and generation of nonverbal cues because all this has to be done in real-time.

An important aspect of HRI design is generating nonverbal behaviors for robots that appropriately accompany speech. This is often inspired by the way humans use nonverbal cues in dialogue. Kanda et al.'s (2007a) robot system automatically generates nonverbal cues, such as nodding and synchronous arm motions, to exhibit its attentional state to the user in correspondence

to the user's arm gestures. Co-speech gesturing—the use of hand, arm, and body gestures that coincide with and match the spoken communication of the robot—has seen good progress in recent years. The best generation methods currently are training on human data. Machine learning methods, often using deep neural networks, are used to learn from hundreds of hours of video recordings of people talking and gesticulating. These methods are now able to generate natural-looking co-speech gestures, with the semantics of gestures— the match between what is being said and what is being expressed by the robot's body—still being a work in progress (Yoon et al., 2022).

Animation framework The most simple and most frequently used approach is to generate motions with an animation framework. That is, a robot designer will typically control each of the joint angles of a robot to set a posture for it; this is called a *key frame*. After the designer prepares multiple key frames, the system interpolates the postures between them and generates smooth motions for the robot.

This requires extensive effort by the designer. Graphical user interfaces (GUIs) are often used to reduce the amount of effort in motion design. The commercial robots Nao and Pepper come with a GUI called Choregraphe, which helps designers visually display the posture of the robot and create desired motions more easily and quickly (see Figure 6.9). Lively is a GUI environment developed by Schoen et al. (2023) to generate lifelike motion for social and other robots that takes obstacles near the robot into account.

Other techniques used for animation or virtual agents can also be used for generating motions for robots. Motion-capture systems can be used to record a timed series of precise human motions, which can then be replicated in robots. Robot designers have also taken advantage of markup languages for virtual agents, such as Behavior Markup Language (BML), in which a designer can specify which gesture an agent should exhibit in combination with speech (Kopp et al., 2006).

Figure 6.9
Choregraphe is a visual editor for the Nao and Pepper robots. It contains a pose editor that allows the robot designer to efficiently generate postures and animations for the robot.
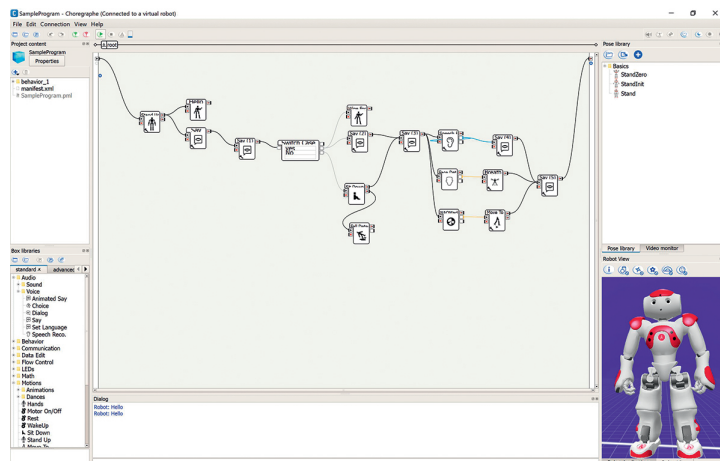
Cognitive mechanisms for robots Another approach to achieving natural behavior in robots is to endow the robot with some form of artificial cognition, which is an artificial equivalent of natural cognition. Although it is, of course, currently impossible to create robots that have the full cognitive capabilities of people, combining artificial equivalents of cognitive mechanisms has proven to be particularly effective for creating lifelike robots. So instead of hand-programming the robot's nonverbal behavior, different cognitive skills are combined, and the resulting behavior often appears natural and lifelike.

Scassellati (2000) developed an embodied cognitive architecture that takes into account salient objects, task constraints, and the attentional state of others to link the robot's perceptions of the world with high-level cognitive skills and related actions, such as joint attention, attribution of intent to others, and social learning. These mechanisms were shown to be particularly effective in the Kismet robot (Figure 2.4). Sugiyama et al. (2007) developed a cognitive mechanism for a robot to replicate human deictic interaction. This involves using pointing (deictic) gestures in reference to a term, such as "this one" or "that one," that signifies a target object the listener can identify. The details of deictic interaction can also depend on the target. For example, we would not point directly at a nearby person because it is impolite. Liu et al. (2013) developed a computational model for a robot that balances two factors, understandability and social appropriateness. It enables a robot to refrain from exhibiting impolite pointing gestures while still keeping its deictic interaction understandable.

6.4 Conclusion

This chapter highlighted the important role of nonverbal cues in communication between humans and robots. The implementation of nonverbal cues into the communicative repertoire of robots still calls for further technical advancement and refinement, particularly because nonverbal cues represent such subtle aspects of communication. Existing research illustrates the relevance of nonverbal communication in HRI while also making clear that much more work needs to be done before robots will be able to act and react in humanlike and natural ways in everyday communication with people.

Questions for you to think about:

- Still not convinced that nonverbal cues are important? Get up right now and have a conversation with someone, but do so without looking at the person's face. How did it go? How did you feel? Also, afterward, ask your communication partner what he or she thought about your behavior and how it made him or her feel.
- Think of a robot use case you are interested in. What aspect of nonverbal behavior is particularly relevant for this scenario? Would gesture or gaze be particularly helpful? How about contingency and

> timing? If you need some inspiration, you can go out and observe people in a similar context and see what they do.
> - Have you ever watched a video where the audio track was a fraction of a second out of sync? Or videoconferenced with someone where the audio lagged? How did that affect the interaction? How long did you think the delay was? What, if anything, did you do to manage the difficulties in the interaction?
> - How would you know if a robot is using is nonverbal cues effectively? Is there a way in which you can measure the quality of nonverbal interaction? Can you measure the outcome of the interaction?

6.5 Exercises

The answers to these questions are available in the Appendix.

* Exercise 6.1 Tipping When waiters and waitresses happen to incidentally touch their customers before they pay for their meal,

1. they get a lower tip.
2. they get a bigger tip.
3. the tip amount doesn't change.

*** Exercise 6.2 Timing Timing is an important component of verbal and nonverbal interaction. Give at least two examples of problems that can emerge in relation to the role of timing in HRI. How might you solve these problems, taking into account how they are solved in other social interactions where the interactants miss out on certain social cues (e.g., texting, time delays in Skype calls, Zoom calls when you can't look at someone in a group to signify you are talking to them)?

** Exercise 6.3 True or false Decide whether the following statements are *true* or *false*:

1. Nonverbal behavior is mostly conscious.
2. Nonverbal behavior can be used to predict social outcomes, such as whether a couple will stay together.
3. Nonverbal and verbal behavior always align.
4. Nonverbal behavior is not very relevant for human–robot interactions.
5. Mimicry is an unconscious behavior, and imitation is a conscious behavior.
6. Gaze is only important for showing someone that you're listening to them.
7. Proxemics (i.e., social distance) is a part of nonverbal communication.

Future reading:

- Admoni, Henny, and Scassellati, Brian. Social eye gaze in human-robot interaction: A review. *Journal of Human-Robot Interaction*, 6 (1):25–63, 2017. doi: 10.5898/JHRI.6.1.Admoni. URL https://doi.org/10.5898/JHRI.6.1.Admoni

- Breazeal, Cynthia, Kidd, Cory D., Thomaz, Andrea Lockerd, Hoffman, Guy, and Berlin, Matt. Effects of nonverbal communication on efficiency and robustness in human-robot teamwork. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 708–713. Institute of Electrical and Electronics Engineers, Piscataway, NJ, 2005. ISBN 0-7803-8912-3. doi: 10.1109/IROS.2005.1545011. URL https://doi.org/10.1109/IROS.2005.1545011

- Mavridis, Nikolaos. A review of verbal and non-verbal human–robot interactive communication. *Robotics and Autonomous Systems*, 63(1): 22–35, 2015. ISSN 0921-8890. doi: 10.1016/j.robot.2014.09.031. URL https://doi.org/10.1016/j.robot.2014.09.031

- Nehaniv, Chrystopher L., Dautenhahn, Kerstin, Kubacki, Jens, Haegele, Martin, Parlitz, Christopher, and Alami, Rachid. A methodological approach relating the classification of gesture to identification of human intent in the context of human-robot interaction. In *IEEE International Workshop on Robot and Human Interactive Communication*, pages 371–377. Institute of Electrical and Electronics Engineers, Piscataway, NJ, 2005. ISBN 0780392744. doi: 10.1109/ROMAN.2005.1513807. URL https://doi.org/10.1109/ROMAN.2005.1513807

- Sidner, Candace L., Lee, Christopher, Kidd, Cory D., Lesh, Neal, and Rich, Charles. Explorations in engagement for humans and robots. *Artificial Intelligence*, 166(1–2):140–164, 2005. doi: 10.1016/j.artint.2005.03.005. URL https://doi.org/10.1016/j.artint.2005.03.005