HRI Module 3: Communication & Emotion

Course 23RIOE313 – Human-Robot Interaction | Textbook 1: Bartneck et al., Human–Robot Interaction: An Introduction | Sections 6.1–6.3, 7.1, 7.2, 7.4, 8.1, 8.2, 8.4 | Emotion: Chapter 9 (9.1 to 9.4)

Module Map

Syllabus topic	Textbook section
Functions of nonverbal cues	6.1
Types of nonverbal interaction	6.2
Nonverbal interaction in robots	6.3
Human-human verbal interaction	7.1
Speech recognition	7.2
Turn-taking in HRI	7.4
Emotion and interaction	Chapter 9 (Emotion): 9.1, 9.2, 9.5 (see Section 7)
Understanding human emotions	Chapter 9: 9.1.2, 9.3.2, 9.4 (Section 7)
Emotions for robots	Chapter 9: 9.3 and 9.3.1 to 9.3.4 (Section 7)
Supplementary: perception, anthropomorphism, acceptance	8.1, 8.2, 8.4 of the perception chapter (Section 8)

> ***Note:*** *The syllabus topics "Emotion and interaction, Understanding human emotions, Emotions for robots" are covered in Section 7, taken from the supplied Chapter 9 ("Emotion", 9.1 to 9.4, with Figures 9.1 to 9.7). Section 8 keeps the earlier notes from the chapter on how people perceive robots (impression formation, anthropomorphism, trust, acceptance, liking) as supplementary material.*

1. Functions of Nonverbal Cues (6.1)

Nonverbal communication is the unspoken layer of interaction: gaze, facial expression, gesture, posture, touch, sounds and timing. It carries information "between the lines" in a subtler, more indirect way than words.

Key functions

- **Adds a layer to speech:** signals mutual understanding, shared goals and common ground.

- **Communicates thoughts, emotions and attention.**
- **Reveals true attitudes:** many cues are automatic or unconscious, so they are seen as unfiltered and more genuine. Example: greeting a disliked acquaintance warmly while looking away briefly, frowning and avoiding touch (incongruence between verbal and nonverbal cues).
- **Predicts outcomes:** "social sensors" measuring gaze and rhythmicity can predict who will exchange business cards at a conference (Pentland & Heibeck, 2010); thin slices of behaviour predicted which couples would split within six years (Carrere & Gottman, 1999).
- **Importance shown by its absence:** people with autism find reading nonverbal cues difficult; travellers abroad feel bewildered when cues differ.

Why it matters in HRI

- Human nonverbal cues toward a robot reveal **engagement, enjoyment and attitude**, and can guide the robot's behaviour.
- Verbal and nonverbal cues can contradict (e.g., saying they like the robot while appearing tense or anxious).
- Robot-produced cues affect the interaction: awkward if gestures mismatch speech rhythm/meaning, or if the robot ignores people's cues.
- Early HRI focused on speech; now nonverbal cues are seen as a **prerequisite for smooth, successful interaction**.
- **Eye-gaze example:** people expect a robot to turn its head and make eye contact, showing it attends to them. A robot staring straight ahead causes the interaction to break down (the missing "social glue").
- Early robot examples: **Kismet** used leaning forward/pulling back to express affect; **Keepon** used gaze and reactive motion to show attention and affect. Many robots use **joint attention** to signal engagement.

Context and culture

- Cues are only appropriate in context, shaped by social and cultural norms.
- Greeting: handshake (West) vs bowing (Japan); the depth of a bow signals status and hierarchy.
- Continuous eye contact or touch may be natural in one culture and threatening or rude in another.
- Example: UK–Japan collaboration on **culturally competent care robots** (cultural knowledge representation, culturally sensitive planning, culturally appropriate multimodal HRI; Bruno et al., 2017).
- Meeting cultural expectations can be the difference between a successful product and a wasted investment.

2. Types of Nonverbal Interaction (6.2)

Cues occur in several modalities at once (sound, movement, gaze), but it helps to study each channel separately, then combine them as needed.

2.1 Gaze and eye movement

- Signals **interest, understanding, attention** and willingness to follow the conversation.

- Functional uses: handing over objects, directing attention to the next tool, **managing turn-taking** (a speaker looks at the next speaker), soliciting and keeping attention.
- Eye tracking can assess gaze patterns and information processing.
- **Joint attention:** partners attend to the same object/area at the same time. In children it scaffolds learning (e.g., learning words; Yu & Smith, 2013); inability can cause developmental difficulties. In adults it signals deep involvement and supports collaboration. **Timing and synchrony** are key.
- **Pupil dilation:** controlled by the autonomic nervous system; pupils dilate when seeing attractive people, and faces with larger pupils are judged more attractive. It can be used on robots to give an impression of attraction.
- **Gaze in HRI** (models derived from human gaze):
    - Leads people into conversational roles: addressee, bystander, non-participant (Mutlu et al., 2012).
    - Controls who speaks next in multiparty interaction (Mutlu et al., 2009).
    - Mutual gaze and purposeful gaze aversion make a robot seem more intentional and thoughtful (Andrist et al., 2014).
    - Gaze while telling a story improved recall of the story for people who received eye contact (Mutlu et al., 2006).
    - Imai et al. (2003) used joint attention so people know what the robot is talking about; Scassellati (1999) studied it for robots learning from humans; also used with autistic children to teach social skills.

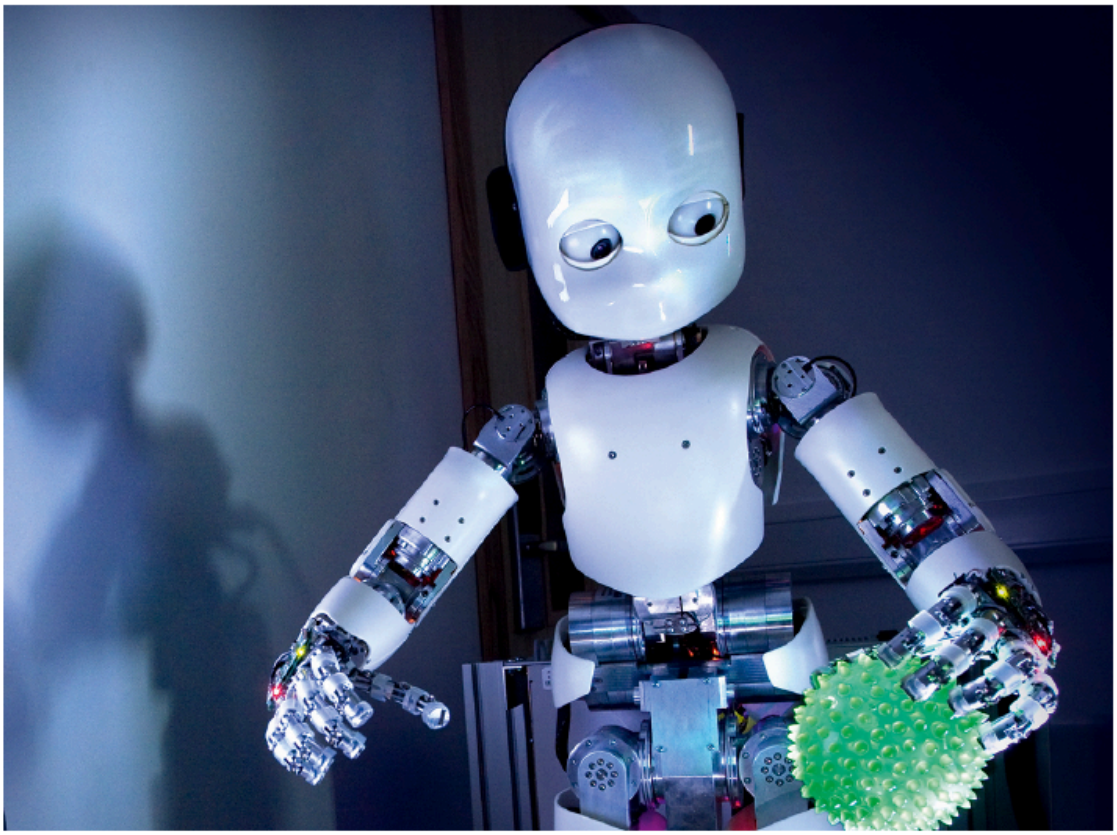


Figure 6.3 (textbook): robot eyes that pitch and yaw let gaze work as a communication channel; iCub attends to the ball in its left hand (joint attention).

2.2 Gesture

Gestures work in place of or along with speech. Categories:

- **Deictic:** pointing at things; establishes joint attention.
- **Iconic:** illustrate speech (arms wide for "a big ball").
- **Symbolic:** carry their own meaning (waving hello/goodbye).
- **Beat:** follow speech rhythm, like conducting an invisible orchestra; also emphasis (hands up while saying "what?").

In HRI: robots gesture with arms, hands, head, ears or tail. Shape, timing, naturalness and smoothness affect perception (Bremner et al., 2009).

- Salem et al. (2013, ASIMO): gestures with speech made the robot seem more anthropomorphic and likable, and people more willing to interact again. **Incongruent gestures** produced even more positive evaluations but hurt task performance. So gestures need careful design and testing.
- **Generation:** co-speech gestures are now mostly produced by **machine learning** trained on hundreds of hours of video of people talking. Gesture dynamics are almost indistinguishable from human ones, but the **semantic fit** to what is said is still imperfect (Yoon et al., 2022).

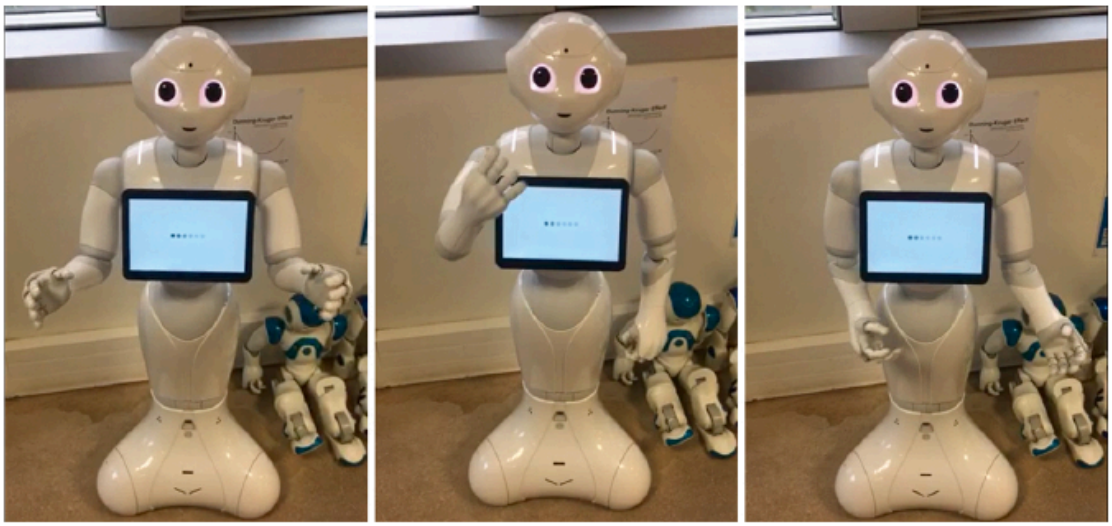


Figure 6.4 (textbook): Pepper using hand gestures to accompany speech; automatically generated beat gestures make its speech look more natural.

2.3 Mimicry and imitation

- **Mimicry** = unconscious replication of another's behaviour; **imitation** = conscious replication (Genschow et al., 2017). Found in primates too.

- Neurological basis: the **mirror neuron system** fires both when observing and performing an action; thought to underlie **facial mimicry**.
- Development: children learn behaviours and norms by imitation; adults blend into surroundings (e.g., gesturing more when speaking Italian), which builds **in-group identity**.
- Social function: signals **liking and positive affect**; being out of sync signals poor communication.
- **Chameleon effect** (Chartrand & Bargh, 1999): subtly mimicking someone's posture/gestures makes them more persuadable. Backfires if too obvious or badly timed (seen as manipulative or insincere).
- **In HRI:**
    - Robot learning by imitation (Argall et al., 2009).
    - Riek et al. (2010): ape-like robot mimicking head gestures improved interaction (though gestures weren't always clear).
    - Wills et al. (2016): a robot mimicking facial expressions with socially contingent head poses received **more donations**.
    - Mimicry can be used to improve interaction and persuade people.

2.4 Touch

- Used deliberately (to calm or console) and incidentally (with people we like). Beneficial mainly within the same social group; appropriateness matters.
- **Midas effect:** waiters get higher tips if they incidentally touch customers (Crusco & Wetzel, 1984).
- Between different social groups results are mixed: may cause negative feelings, though incidental touch reduced indirect prejudice (Seger et al., 2014).
- **Challenges for robots:** safety (hugging, holding, handshakes), pinch points (e.g., Nao and Pepper under arms and near hips), fragility under human-level handling forces. Touch is often an afterthought; only a few robots are designed for it (e.g., **Telenoid**, designed to be hugged).
- **Findings:**
    - Animal-like robots (**Paro**, **Haptic Creature**) can reduce stress and anxiety when people initiate touch.
    - Chen et al. (2014): people accepted being touched by a nursing robot; **functional touch** (cleaning the arm) was rated better than **affective touch** (comforting).
    - Wullenkord et al. (2016, Nao): touch generally improved attitudes, but people with initially negative feelings about robots became **more negative** after touching.
- Touch is integral to functional tasks (object handover, manipulation) and social tasks (handshake greeting).

2.5 Posture and movement

- Posture and movement signal emotional state: slow movement, drooping shoulders and lethargy = downcast; fast movement and upright bearing = positive. Interpretable even from **point-light displays** (Alaerts et al., 2011).
- Posture signals attention, engagement and attraction: crossed arms = defensive, open arms = invitation; knees toward each other = engagement, turning away = wish to end the interaction.

- **Robots:** body posture is the main route for emotion when a robot lacks facial features. Beck et al. (2010): affective postures (Nao) improved understanding of emotional state. Xu et al. (2014): people interpreted robot postures and even **adopted the emotions** they perceived.
- **Micromovements** (small random perturbations of actuators) create a lifelike impression; velocity or amplitude of motion can signal excitement (Belpaeme et al., 2012). Used on pet-like robots.
- Example: **Thrifty Faucet** uses posture to show 15 lifelike motion patterns (seeking, curiosity, rejection) to encourage sustainable water use.

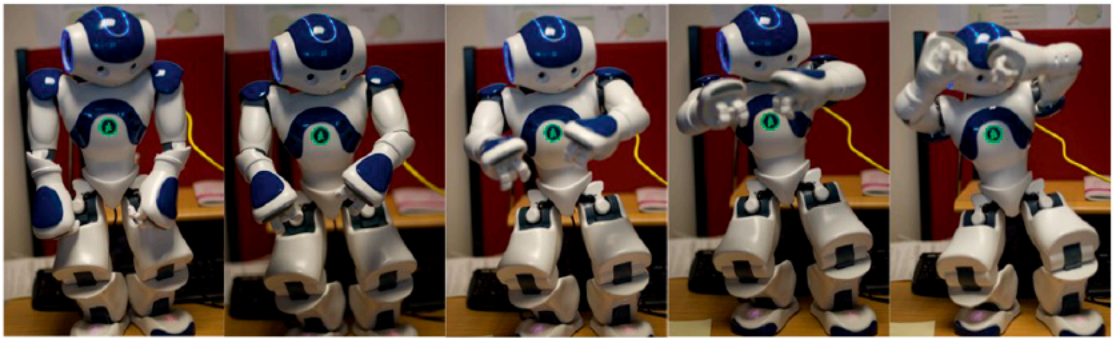


Figure 6.8 (textbook): Nao using body postures to express emotion, morphing between sad (left) and fearful (right) (Beck et al., 2010).

2.6 Interaction rhythm and timing

- Timing supports **turn-taking**; gaze and gesture can guide attention or signal the end of a turn.
- **Rhythmicity/synchrony:** partners become "rhythmically entrained", acting to the same beat, like dancing (Warner et al., 1987). Being out of sync suggests missed signals and leads to worse outcomes and less positive attitudes.
- HRI evidence: a robot entrained with the human is rated more **lifelike**, and people interact longer with an entrained dancing robot (Michalowski et al., 2007). Rhythm aids anticipation in team collaboration (Hoffman & Breazeal, 2007). Rhythmic music improved human–robot surgical team performance (Siu et al., 2010).

3. Nonverbal Interaction in Robots (6.3)

3.1 Robot perception of nonverbal cues

- Uses standard **pattern recognition**. Posture and gesture recognition are well studied; sensors include cameras, depth cameras and user-worn sensors recording time series data.
- **Machine learning pipeline:** collect a database (thousands to millions of data points, e.g., people waving, pointing, beckoning) → **label** every data point → train a **classifier** iteratively → deploy on the robot (Mitra & Acharya, 2007).
- **Engagement detection:** in HRI users sometimes don't attend to the robot, so perceiving engagement is crucial. Rich et al. (2010) combined eye contact and back-channeling; Sanghvi et al. (2011) used affective posture and body motion with a game companion robot.

- Extra equipment: eye trackers, motion capture. **Tactile sensing:** piezoelectric polymer sensors in silicone rubber (Taichi et al., 2006); flexible capacitive sensors or skin with nano hairs (Yogeswaran et al., 2015).

3.2 Generating nonverbal cues in robots

- Not trivial: cues must be **contingent** (user snaps fingers → robot blinks immediately), **coordinated** with each other and with speech (meaning and timing), and done in **real time**.
- Kanda et al. (2007a): robot automatically generates nodding and synchronous arm motions to show attention in response to the user's gestures.
- **Co-speech gesturing** is best done by deep neural networks trained on human video; semantics still a work in progress.

Approaches to generating motion:

1. **Animation framework (most common):** the designer sets joint angles to create a posture, called a **key frame**; the system interpolates between key frames for smooth motion. Labour-intensive, so GUIs help: **Choregraphe** (Nao/Pepper) with a pose editor; **Lively** (Schoen et al., 2023) generates lifelike motion that accounts for nearby obstacles.
2. **Motion capture:** record precise human motions and replay on the robot.
3. **Markup languages:** e.g., **Behavior Markup Language (BML)** to specify which gesture accompanies speech (Kopp et al., 2006).
4. **Cognitive mechanisms:** combine artificial cognitive skills so natural behaviour emerges rather than being hand-programmed.
    - Scassellati (2000): embodied cognitive architecture using salient objects, task constraints and others' attention, enabling joint attention, intent attribution and social learning (effective in **Kismet**).
    - Sugiyama et al. (2007): robot replicates **deictic** interaction ("this one"/"that one").
    - Liu et al. (2013): model balancing **understandability** and **social appropriateness**, e.g., avoiding impolite direct pointing at a nearby person.

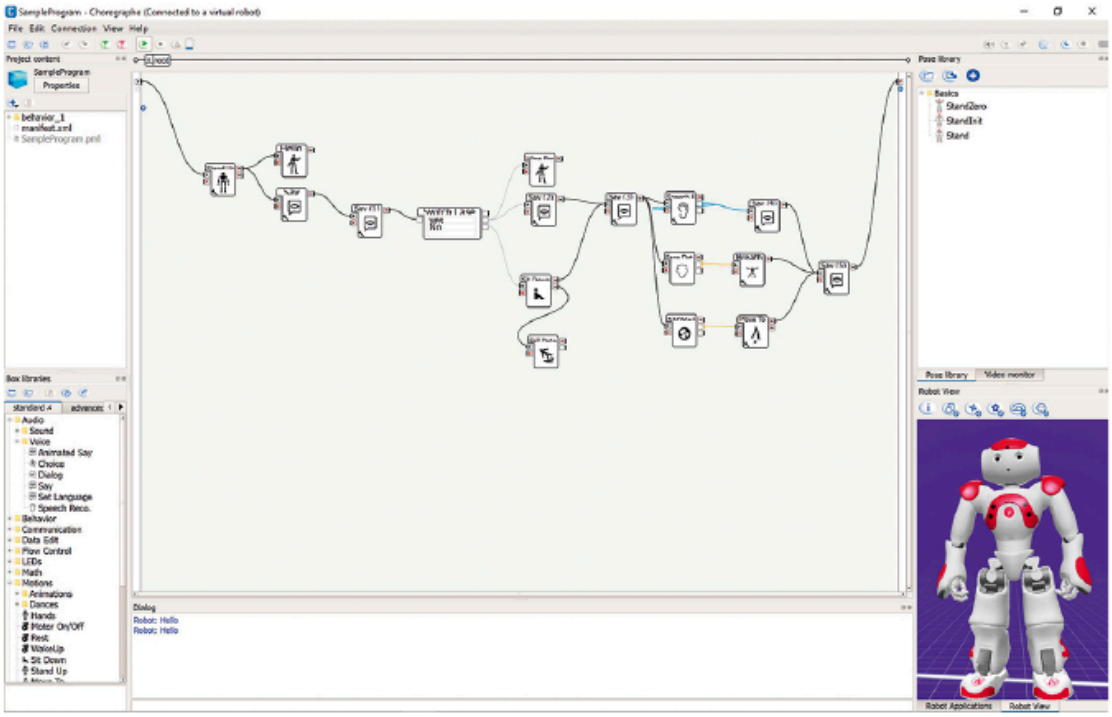


Figure 6.9 (textbook): Choregraphe, the visual editor for Nao and Pepper, with a pose editor to create postures and animations (key frames).

Conclusion: nonverbal cues are vital but subtle, and technical refinement is still needed before robots act humanlike in everyday communication.

4. Human–Human Verbal Interaction (7.1)

- Speech conveys information and also creates **joint attention and shared reality**. It is complex and ambiguous: shifting emphasis in *"She said she did not take his money"* changes the meaning (belief, disbelief, accusation).
- Enriched by **paralinguistic** information: prosody, plus gaze, gestures and facial expressions.

4.1 Components of speech

- **Utterance:** smallest unit of spoken language; often shorter and less grammatical than written sentences, with pauses and disfluencies ("you know, I was, like, uhm...").
- **Words:** smallest units that convey meaning.
- **Phonemes:** small units of sound that make up words; *pat* = [p][a][t]; changing [p] to [b] gives *bat*.
- **Conversational fillers:** (e.g., "uh-huh") keep the conversation going and let listeners signal attention, understanding, surprise or shared emotion without interrupting. They raise efficiency and the sense of shared reality.

4.2 Written text vs spoken language

Written	Spoken
Strict grammar and syntax	More liberal, informal
Unidirectional; cannot be adjusted mid-delivery, so needs precision	Misunderstandings clarified on the fly; speakers adapt when the listener doesn't understand

For HRI, natural smooth communication requires: transcribing speech into words, understanding words and choosing responses, and generating spoken language, all from verbal speech, which is harder than text.

5. Speech Recognition (7.2)

Speech recognition = **ASR** (automatic speech recognition) or **STT** (speech-to-text): converts a recording of speech into text. It does **not** understand meaning.

Why HRI is hard

- ASR was developed for dictation and device control, assuming a quiet room and a directional microphone on the speaker.
- In HRI the speaker is at a distance, other people and environmental sounds are picked up, and even the robot's own **mechanical noise** is recorded. Microphone arrays and signal processing help, but many robots lack them.
- Workaround: **close miking** (lapel microphone or headset).
- **Multimodal perception:** humans also use vision (**McGurk effect**: the same sound is heard as "baa" or "faa" depending on mouth shape) and context knowledge (peach vs beach).

Basic principles

1. Digital recording in the **time domain** (amplitude per time step, e.g., every 1/16,000 s).
2. Convert to the **frequency domain**, where phonemes have distinct signatures (an *o* differs from an *a*).
3. Transcribe using an engine.

- **Before ~2010:** rule sets, **Gaussian mixture models** and **hidden Markov models**, i.e., probabilistic models of how phonemes and words combine ("robot" more likely than "lobot").
- **Now: deep learning** (large neural networks, sequence-to-sequence models) that does this implicitly.
- Performance: handles noise, crowds, ill-formed speech, children, accents. Word error rate (WER) of **1.4%** in one test vs **5.8%** for human transcribers (Zhang et al., 2022). Mostly **cloud-based**; onboard ASR is typically poorer.

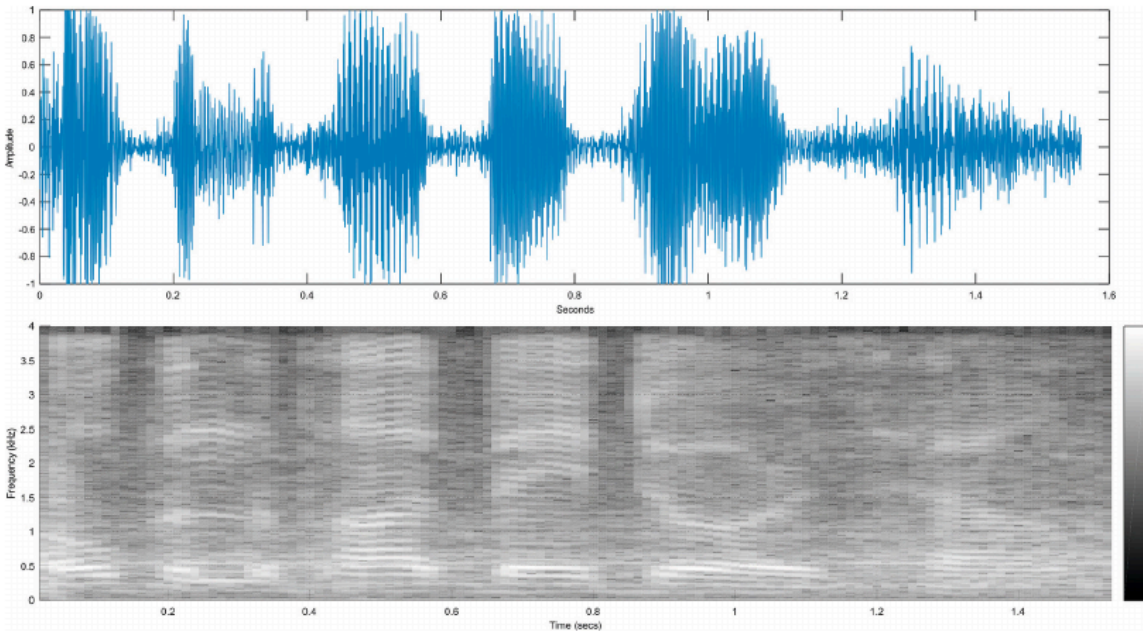


Figure 7.2 (textbook): the sample "Open the pod bay doors, HAL" in the time domain (top, amplitude) and frequency domain (bottom). ASR must transform this data into text.

Limitations

- Atypical speech, insufficiently trained speakers, dialects and non-native speakers.
- Noisy, reverberating or crowded spaces.
- **Proper nouns** (e.g., *Margaret*, *Launceston Street*).
- **Constraining** the vocabulary (digits, simple commands) improves accuracy.
- Progress is rapid, e.g., **OpenAI Whisper** learned from hundreds of thousands of hours of unlabelled speech.

Practice in HRI

- **Cloud services:** send recorded speech, get text back; best performance, frees robot computing. Offered by Baidu, Google, IBM, Microsoft, Nuance, OpenAI (costs on the order of 1 cent per recognition). Open source: Mozilla Common Voice and DeepSpeech.
- **Onboard ASR** when there is no reliable internet; lower performance.
- Engines have simple **APIs**, return a **confidence value** and sometimes alternative transcriptions with confidences.
- **ROILA** (RObot Interaction LAnguage): an artificial language designed so words sound distinct, easing ASR (e.g., "go forward" = *kanek koloke*; "go back" = *kanek nole*). Other constructed languages: engineered (Loglan, ROILA), auxiliary (Esperanto), artistic (Klingon).
- **Voice-activity detection (VAD):** distinguishes speech from silence and other sounds (e.g., music vs talking). In HRI it gives the impression the robot is listening and enables turn-taking without recognition. Deep learning improved VAD; **OpenSmile** is a leading free package. Combined with **sound-source localisation** (multiple microphones) the robot can look at the speaker.

6. Turn-Taking in HRI (7.4)

Back-channeling

- Listener responses showing engagement: "really?", "uh-huh", plus nonverbal nods and smiles.
- Personal assistants use visual signals (pulsing light). Robots can use **verbal** back-channels (nonlexical "uh-huh"/"hmm", phrasal "yeah", "tell me more") and **augment** them with blinking lights or a gentle hum.
- Challenge: **when** to back-channel; it depends on the speaker's verbal and nonverbal cues. Park et al. (2017a): children preferred a robot using a back-channel **prediction model** with contingent signals.

The role of timing

- Delayed responses are seen as disturbing; very quick responses can seem insincere (Sacks et al., 1974; Heldner & Edlund, 2010).
- Timing depends on cognitive load: yes/no answers are faster than fully formed replies. In telephone conversations, "yes" answers average **~100 ms**, while responses to undesired offers take almost **500 ms** (Strömbergsson et al., 2013). People often anticipate and answer before a question ends.
- **Robots are much slower:** the sequential processing chain (ASR → understanding → dialogue management → speech synthesis) can take several seconds.
- **Mitigations:** conversational fillers (Shiwa et al., 2008) or visual signals showing the robot is "thinking"; but these are poor substitutes for fast turn-taking.
- **Promising techniques: just-in-time speech synthesis** (start speaking before the full sentence is planned) and **incremental spoken-dialogue processing** (act on instructions before they finish; Baumann & Schlangen, 2012).

7. Emotion and Interaction (Chapter 9: Emotion, 9.1 to 9.4)

Covers the syllabus topics **Emotion and interaction**, **Understanding human emotions** and **Emotions for robots**. Social robots are often designed to **interpret** human emotion, **express** emotion, and sometimes even have **synthetic emotion** driving their behaviour. Emotions are not in every social robot, but taking them into account improves the intuitiveness of HRI.

7.1 What are emotions, mood and affect? (9.1)

- **Evolutionary role:** emotions help individuals respond to environmental factors that promote or threaten survival. They **prepare the body** for behavioural responses, help in **decision-making** and **facilitate interpersonal interaction**.
- **Appraisal:** emotions arise as an appraisal of situations and prepare us for a response (Gross, 2007; Lazarus, 1991).
    - Anger (someone pushes in line): adrenaline makes us ready for action and our expression signals the other person crossed a line.
    - Sadness (not invited to a party): hampers quick action, makes us reconsider our behaviour, and evokes empathetic responses from others. So emotions also **modulate the behaviour of others**.

- **Affect:** the comprehensive term for the entire spectrum of emotionally laden responses, from quick subconscious reactions to an external event to lingering complex moods such as love. Within affect, a distinction is made between **emotions** and **moods**.

Aspect	Emotion	Mood
Cause	Identifiable source (an event, or seeing emotion in others)	Often no clear cause; result of environmental, incidental and cognitive processes
Object	Directed at a specific object or person; often externalised	Diffuse and internal; lacks a clear object
Duration	Shorter-lived	Longer-lasting
Example	Happiness on getting a promotion; anger when the phone battery dies during an important call; jealousy when a colleague gets a company car	Apprehensive mood while waiting a week for medical test results; the warm feeling of a sunny week with friends

Emotion and interaction (9.1.1)

- Emotions are not just internal; they are a **universal communication channel** that has helped communicate internal affective states to others and was likely important for human survival.
- Emotions give the outside world information in **two ways**:
    - About **you and your possible future actions** (anger and frustration signal preparation for an aggressive response).
    - About **the environment** (an expression of fear alerts others to a fast-approaching grizzly bear before you can scream).
- Either way emotion gives others an **incentive to act**: with anger someone may step down and try to defuse the situation; with fear people scan the environment for a threat (Keltner and Kring, 1998).
- Successful communication of emotion **promotes survival, enhances social bonds, and minimises the chance of social rejection and physical aggression** (Andersen and Guerrero, 1998).

Conceptualising human emotions (9.1.2)

- Aristotle listed **14** emotions (anger, love, mildness...); Ekman (1999) listed **15** basic emotions (pride in achievement, relief, satisfaction, sensory pleasure, shame...). A definitive list is impossible: emotions vary between people and cultures, language does not map perfectly to emotions, and some emotions overlap.
- Anger, sadness and happiness are likely candidates for **core emotions**.

- **Ekman and Friesen (1975):** six basic **facial expressions** recognised across cultures. These are often mistaken for a set of basic emotions we experience; they were only meant as a basic set of emotions we **express via our faces** and that different cultures recognise.

	Primary (basic) emotions	**Secondary (reactive) emotions**
Nature	Quick, gut-level responses	Reactive and reflective
Culture	Considered universal across cultures	Differ across cultures
Examples	Amusement, anger, surprise, disgust, fear	Pride, remorse, guilt

- **Debate:** no consensus on which emotions belong in each category (Holm, 1999; Greenberg, 2008), and some scholars (Ortony and Turner, 1990) argue basic emotions do not exist at all.
- **Dimensional view, Russell (1980) circumplex model:** emotions are cognitive interpretations of sensations arising from two independent neurophysiological systems, **arousal** and **valence**. So emotions lie on a **two-dimensional continuum** instead of discrete independent basic emotions (Posner et al., 2005). Widely studied and confirmed across languages and cultures.
- Limitation: a meta-analysis (Remington et al., 2000) found it is a reasonable representation of self-reported affect, but not all affective states fall into the predicted regions and some cannot be consistently placed.

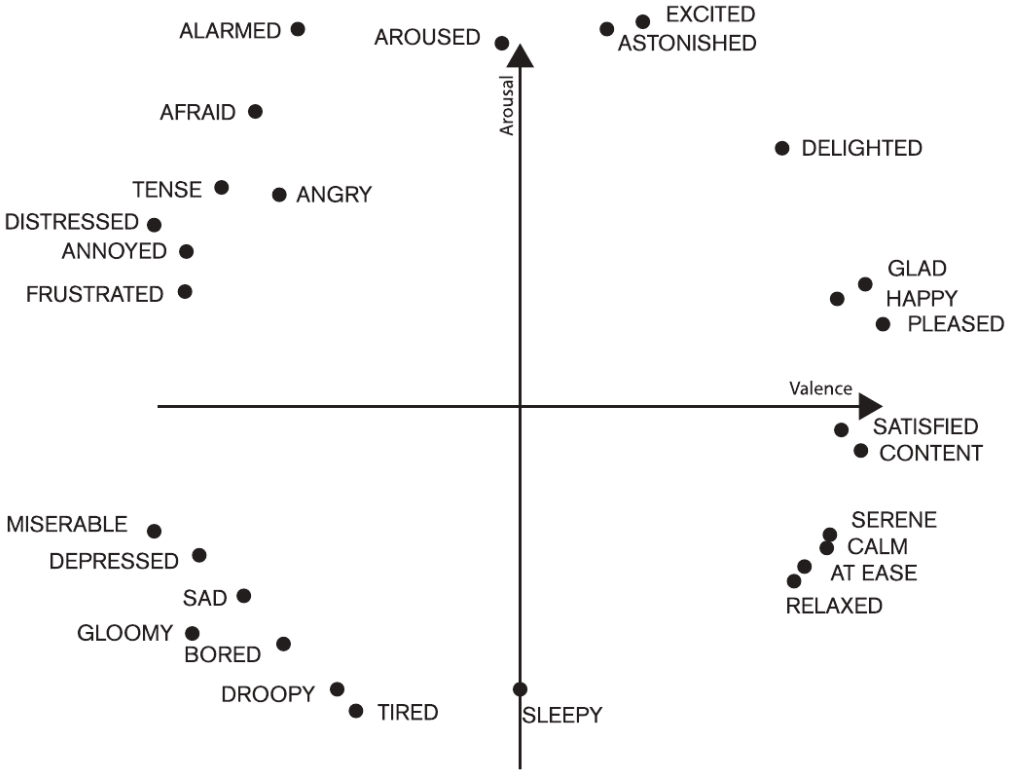


Figure 9.1 (textbook): Russell's (1980) circumplex model of affect. Vertical axis = arousal; horizontal axis = valence. For example ALARMED, AFRAID, ANGRY, TENSE sit at high arousal and negative valence; EXCITED, DELIGHTED at high arousal and positive valence; SERENE, CALM, RELAXED at low arousal and positive valence; DEPRESSED, SAD, BORED, TIRED at low arousal and negative valence.

Facial feedback hypothesis: expressing emotion also informs *you* how you feel. Facial movement influences experienced emotion. Participants holding a pencil between their **teeth** (forced smile) rated a comic as slightly funnier than those holding it in the hand; holding it between the **lips** made it less funny (Strack et al., 1988). Botox (which paralyses facial muscles) reduces the intensity of experienced emotions (Davis et al., 2010).

7.2 When emotions go wrong (9.2)

- The importance of emotion shows when a partner **fails to understand** the other's emotion or **fails to respond appropriately**. Even tiny glitches have consequences: e.g. mistaking **sarcasm** for a genuine response causes misunderstanding and hurt feelings. It is worse when someone is consistently unable to perceive, express or respond to affect.
- **Depression:** problems with emotional responsiveness are a defining symptom (Joormann and Gotlib, 2010). Depressed individuals can understand others' feelings and express their own, but show a **reduced emotional response to positive stimuli** (e.g. rewards) and have recurring negative thoughts about past, present and future. This leads to social isolation and more loneliness, worsening their state.
- **Autism spectrum disorders:** may find it difficult to correctly **interpret** displays of emotion, so they cannot intuitively understand partners' needs and often respond inappropriately.
- **Impaired expression:** e.g. facial muscles impaired after a **stroke**, making it hard for partners to infer internal states.
- An inability to express and interpret emotions impairs effective, smooth interaction with people. Likewise, **interaction with robots is difficult if the robot cannot express and interpret emotional states**.



Figure 9.2 (textbook): Kaspar (2009 to present), a "minimally expressive" robot built using brackets, servo motors and a surgical silicone mask, used in autism therapy. (Source: Kerstin Dautenhahn, Ben Robins, Adaptive Systems Research Group, University of Hertfordshire, UK)

7.3 Emotions for robots (9.3)

- Emotion is an important **communication channel** in HRI. When a robot expresses emotion, people ascribe a level of **social agency** to it (Breazeal, 2004a; Novikova and Watts, 2015).
- Even a robot **not designed** to express emotion may be interpreted as if motivated by emotional states. A robot not programmed to share, understand or express emotion risks being seen as **disinterested, cold or rude**.
- Designers must consider: what emotions the design and behaviour **convey**, whether and how the robot will **interpret** emotional input, and how it will **respond**.

Emotion interaction strategies: mimicry (9.3.1)

- The most straightforward way to program emotional responsiveness is **mimicry**. In humans, mimicking creates an idea of **shared reality** (you understand the other's situation), which creates **closeness** (Stel et al., 2008).
- **Exception: anger.** Yelling back at an angry person usually does not help mutual understanding or resolve the conflict.
- For a robot, mimicry only requires **recognising** the human's emotion and **reflecting it back**, which avoids the harder task of formulating an appropriate response. Humans also basically expect it: we may excuse friends for not knowing how to cheer us up, but expect them to respond to our sadness by lowering brows and head and becoming more soft-spoken.
- **Expectation management:** if users see the robot as emotionally responsive they extend expectations to other social norms. A user upset about a confrontational meeting may expect the robot to ask about it; if it merely says "have a great day at work!" he is disappointed. **The robot's emotional responsiveness should match its capability to meet other expectations.**

Artificial perception of emotions: understanding human emotions (9.3.2)

- Robots must register a wide variety of emotional cues, explicit and subtle. E.g. a robot that reacts to aggression such as an item thrown at it needs both **human behaviour recognition** and **object recognition**.
- **Facial cues (computer vision):** the most typical approach. Given a data set of labelled (frontal) faces, machine learning (e.g. **deep learning**) extracts features to recognise a range of facial emotions. Famous example: **smile detection** in digital cameras. Affect recognition can also use other visual cues such as **walking patterns**, removing the need for a clear view of the face (Venture et al., 2014).

Smile detection in cameras: *if a group poses, the camera only takes the shot when everyone smiles, partly replacing the timer, which could never guarantee that everybody looked at the camera and smiled.*

- **Speech (prosody):** the second-most-important channel. Prosody = patterns of stress and intonation. **Happy** speakers tend to talk with a **higher pitch**; **sad** speakers speak **slowly with a lower pitch**. Pattern recognition (machine learning) is used to infer emotion from speech (El Ayadi et al., 2011; Han et al., 2014).

- **Other modalities: skin conductance** changes with affective state (used in the polygraph/lie detector), but sensors have had only **limited success in HRI** (Bethel et al., 2007).

Expressing emotions with robots (9.3.3)

- Typically via **facial expressions**, mimicking how people display emotion, a good example of using human behaviour studies to design robot behaviour.
- **Ekman's Facial Action Coding System (FACS):** human facial muscles are grouped as **action units (AUs)** and emotions are described as combinations of AUs (Ekman and Friesen, 1978). E.g. a **happy face** (smile) uses **AU6** (orbicularis oculi / pars orbitalis, raises the cheek) and **AU12** (zygomaticus major, raises mouth corners).
- Robots use simplified equivalents of facial muscles. Hashimoto et al. (2013): a face with **soft rubber skin and 19 pneumatic actuators** uses AUs (AU6 and AU12 for happiness). Others using simplified human facial cues: **Kismet, Eddie, iCat, eMuu** (and Flobi).



Figure 9.3 (textbook): Emotions expressed through mechanical facial expressions. Left: eMuu (2001). Middle: iCat (2005 to 2012). Right: Flobi (2010). (Source: Left and middle, Christoph Bartneck; right, University of Bielefeld)

- **Other modalities:** body movements and prosody. Even **non-anthropomorphic** robots can express affect through **navigation trajectories/motion patterns** (a cleaning robot: Saerbeck and Bartneck, 2010; a flying robot: Sharma et al., 2013), and through **speed of motion, body posture, sound, colour and orientation** towards the person (Bethel and Murphy, 2008).

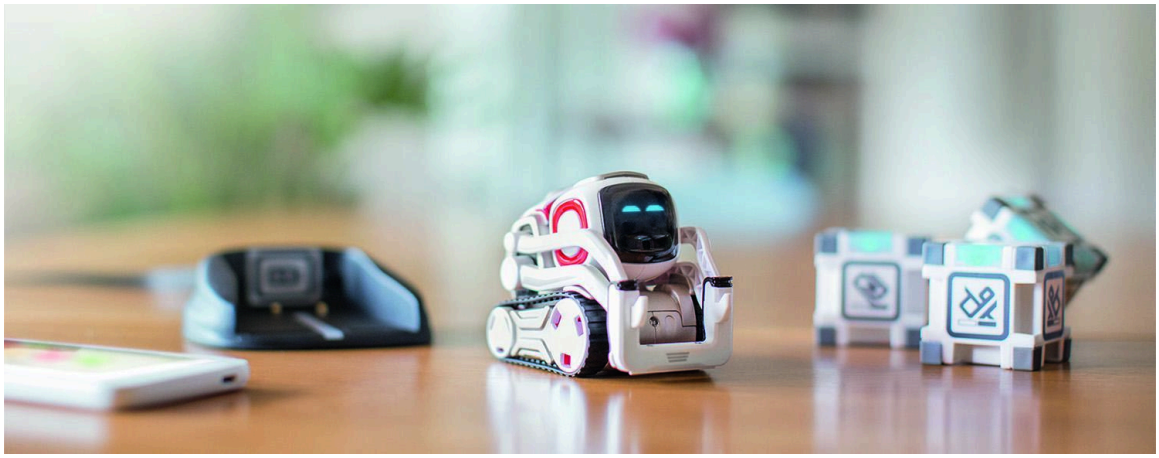


Figure 9.4 (textbook): Non-anthropomorphic robots can express emotion through their behaviour or through added expressive features such as lights. Anki, the producer of Cozmo (2016 to 2019), describes its robot as having "his own lively personality, driven by powerful A.I., and brought to life with complex facial expressions, a host of emotions and his own emotive language and soundtrack." (Source: Anki)

Emotion models (9.3.4)

- Psychologists have captured emotions in **formal models** (Plutchik and Conte, 1997; Scherer, 1984). Benefit: emotions become a **numerical representation**, which suits computers and robots, and models put emotional categories **in relation to each other** (e.g. happiness as the polar opposite of sadness, or a distance function between emotions).
- Models can capture the **user's** emotional state and also the **robot's own** emotional state to **drive its behaviour**. Example: a robot with an almost empty battery acts tired and announces it needs rest; once on the charger it updates its internal state to happy. Expressing this gives the user access to the robot's internal state and enriches interaction.

(a) OCC model (Ortony, Clore and Collins, 1988)

- Classic model used in some robots; named after the authors' initials.
- Specifies **22 emotion categories** based on **valenced reactions** to: **consequences of events**, **actions of agents** (including oneself), and **aspects of objects** (attractive or unattractive).
- Also structures the **variables** that set the intensity of each emotion type (e.g. likelihood of an event, familiarity of an object). Detailed enough to cover most situations an emotional robot would face.
- Problem: most robots cannot express all 22 emotions, and implementing 22 is challenging, so designers often reduce categories, commonly to **Ekman's six basic facial expressions** (reliably recognised across cultures; Ekman, 1992). But a robot with only six emotions gives a **limited interaction experience**.

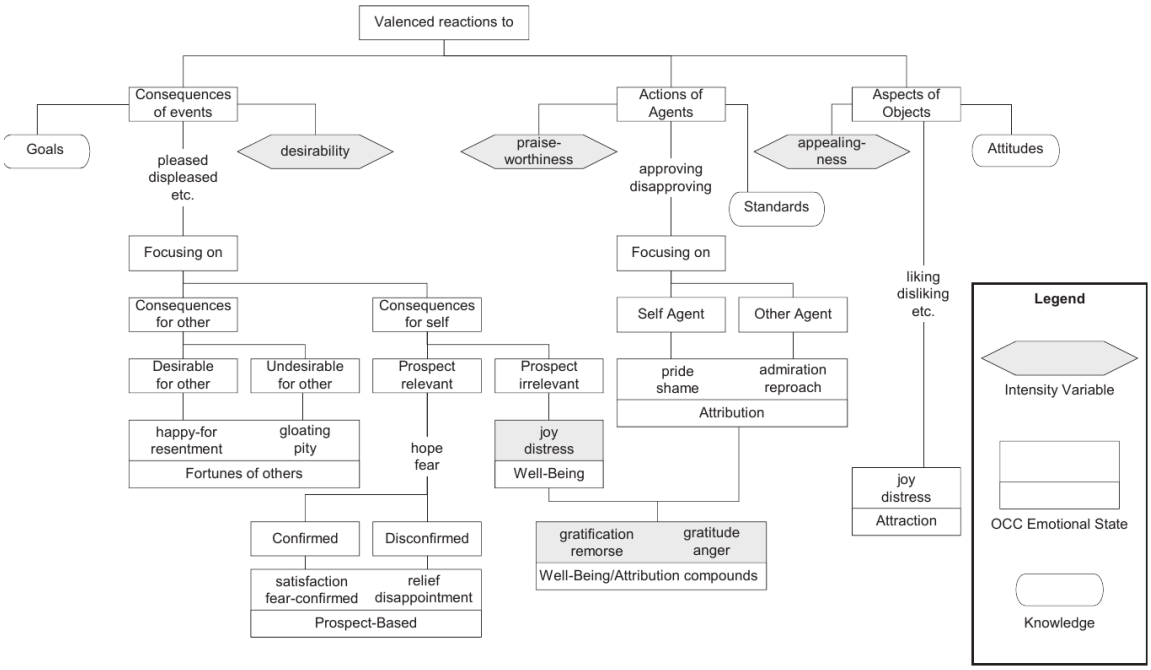


Figure 9.5 (textbook): The OCC model of emotions. Valenced reactions to consequences of events (goals, desirability), actions of agents (standards, praiseworthiness) and aspects of objects (attitudes, appealingness), branching into pairs such as joy/distress, hope/fear, happy-for/resentment, gloating/pity, pride/shame, admiration/reproach, gratification/remorse, gratitude/anger. Legend: intensity variable, OCC emotional state, knowledge.

(b) Dimensional models (point in a multidimensional space)

- Perhaps more popular than OCC. **Russell's 2D space of arousal and valence** (Figure 9.1) captures a wide range of emotions on a plane and is one of the **simplest models with enough expressive power for HRI**.
- Weakness: the original 2D circumplex places **"angry" and "afraid" side by side**, although most people see them as vastly different.
- Fix: a **third axis** gives Mehrabian's (1980; Mehrabian and Russell, 1974) **3D continuous space: Pleasure (P), Arousal (A), Dominance (D)**, the **PAD model**. Used on many social robots including **Kismet**, to model both the user's and the robot's emotional state (Breazeal, 2003).

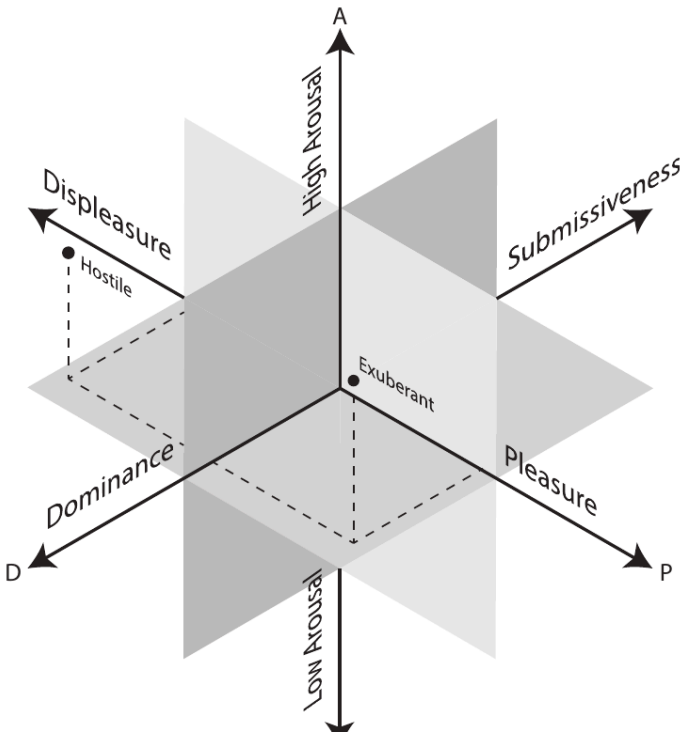


Figure 9.6 (textbook): The PAD emotion model. An emotion is a point in a 3D space with axes pleasure (P), dominance (D) and arousal (A). Examples shown: "Hostile" (displeasure, dominance, high arousal) and "Exuberant" (pleasure, high arousal).

Model	**Type**	**Key idea**	**Used for**
Ekman six expressions	Discrete categories	Six facial expressions recognised across cultures	Simple robot faces; limited range
OCC (1988)	Cognitive appraisal, 22 categories	Valenced reactions to events, agents' actions, objects; intensity variables	Rich robot emotion reasoning; complex to implement
Russell circumplex (1980)	2D continuous	Arousal and valence	Simple but expressive enough for HRI; angry and afraid too close
PAD (Mehrabian)	3D continuous	Pleasure, Arousal, Dominance	User and robot emotion state (e.g. Kismet)

7.4 Challenges in affective HRI (9.4)

Despite considerable work on perception, representation and expression of emotion in virtual agents and robots, open challenges remain:

- **Faces alone are not enough:** it is virtually impossible to correctly read emotions from facial information alone, and people struggle with still faces, so robots will too. Adding **context of the interaction**, **animated rather than still expressions** and **body language** raises recognition rates for both people and algorithms.



Figure 9.7 (textbook): Can you tell if the tennis player just scored or lost a point? People struggled to read strong emotions from static faces alone, but could when seeing only the body posture (Aviezer et al., 2012). (Source: Steven Pisano)

- **Acted training data:** almost all algorithms are trained on emotions **acted out by actors**, so they are exaggerated and unlike everyday emotion. Software only recognises emotions shown with a certain **exaggerated intensity**; accuracy on **subtle** expressions drops dramatically (Pantic et al., 2007; Bartneck and Reichenbach, 2005), limiting real-world use.
- **Limited output categories:** most software returns probabilities for only Ekman's **six basic expressions** or a point in a 2D/3D space, missing emotions like **pride, embarrassment, guilt or annoyance**.
- **Individual differences:** people express universal emotions with different intensity, in different contexts and with different meanings. Interpreting emotion needs sensitivity to **individual affective quirks**, which humans learn through years of experience (you can tell your partner is laughing out of annoyance, a new acquaintance cannot). Robots decode from **momentary snapshots** of the face and do **not build long-term models** of a partner's affect, emotion and mood.
- **Fooling users:** emotional responsiveness can make users think the robot **genuinely experiences emotions**. A robot only **displays** emotional states from a computational model; it does not replace the visceral experience. **Affective cognition**, a full socio-emotional repertoire expressed and recognised across users and contexts, remains elusive.

Conclusion (9.5)

- Emotions have **intrapersonal functions** (evaluating the situation, motivating action) and an **interpersonal function** (informing others of our mental state and, by extension, expected behaviour).

- For smooth human-robot interaction, a robot must both **recognise** the human's emotions and **generate its own emotions** to inform the human of its internal state.

Quick comparison: what to remember

Topic	Key point
Affect / emotion / mood	Affect is the umbrella term; emotions have a specific cause and object and are short-lived; moods are diffuse and longer-lasting
Functions	Survival, decision-making, preparing a response, communicating internal states and the environment, social bonding
Basic vs secondary	Basic: amusement, anger, surprise, disgust, fear (universal, gut-level). Secondary: pride, remorse, guilt (culture-dependent, reflective)
Russell vs PAD	2D (arousal, valence) vs 3D (pleasure, arousal, dominance); PAD separates angry from afraid
Perception channels	Face (computer vision), body/gait, speech prosody, skin conductance (limited success)
Expression channels	Facial AUs (FACS: AU6 + AU12 = smile), body movement, prosody, motion trajectories, light, sound, colour
Main challenges	Context, acted data, subtle emotion, limited categories, individual differences, no long-term models, false impression of genuine emotion

8. How People Perceive Robots: Impressions, Anthropomorphism and Acceptance (8.1, 8.2, 8.4)

Covers how people form impressions of robots, anthropomorphise them, and the consequences (trust, acceptance, liking). This follows the earlier emotion-and-perception content of the notes, kept as supplementary material.

8.1 Impression formation (8.1)

- People form impressions **quickly and automatically** (snap judgments within milliseconds).
- **Dual-process models:** two ways of processing information:
    - **Automatic** (intuitive, quick, may be outside awareness; System 1, associative).
    - **Deliberate** (conscious, slow, effortful; System 2, rule-based/controlled).

- Automatic responses come first and can be adjusted by deliberate processing. Models differ: **sequential** (Evans & Stanovich, 2013) vs **parallel** (Smith & DeCoster, 2000). Deliberate processing only occurs with motivation and capacity, and is not perfectly rational.
- Automatic associations come from **prior experience**: sci-fi villains (Terminator) lead to negative first reactions; friendly robots (Doraemon, Astro Boy) to positive ones.
- Implicit and explicit attitudes can differ: people may be more negative about robots on implicit measures than they say (de Graaf et al., 2016).
- People also attribute human features (intentions, emotions, mind: **agency** and **experience**, Gray et al., 2007) to robots: **anthropomorphization**, which can also be viewed through a dual-process model.

8.2 Anthropomorphism (8.2)

- **Anthropomorphism** = perceiving/judging a humanlike form; **anthropomorphization** = attributing humanlike characteristics to nonhuman entities.
- Early HRI equated it with **humanlike appearance** (engineering view).

Psychological anthropomorphism (Epley et al., 2007): three determinants

1. **Effectance motivation:** desire to explain and understand others' behaviour and regain control in novel situations; anthropomorphizing can reduce stress and anxiety. Explains why robot movement is read as a social cue (Erel et al., 2019).
2. **Sociality motivation:** lonely people (situational or chronic) anthropomorphize robots more (Eyssel & Reich, 2013).
3. **Elicited agent knowledge:** using commonsense knowledge of social actors; e.g., people explained dating norms more to a male robot (Powers et al., 2005). Useful for design.

People use self-related (**anthropocentric**) knowledge to make sense of nonhuman things; humanlike appearance and behaviour trigger it, affecting perception and behaviour toward the robot.

The process of anthropomorphization

- **Media equation** (Reeves & Nass, 1996): people treat media/computers as social actors (the **CASA** approach: computers as social actors), supported by research on social robots; judgments are automatic.
- **Złotowski et al. (2018):** System 1 snap judgment ("humanlike or not?") then System 2 correction; outcomes are **implicit vs explicit anthropomorphism**.
- **Lemaignan et al. (2014a) phases:** pre-initial (a priori impression) → initialization (revise/extend) → **familiarization** (more realistic view; anthropomorphic inferences likely decrease) → stabilization (comprehensive judgment).
- **Lemaignan et al. (2014b) three-stage model:** I automatic evaluation (no actual interaction needed); II interaction creates a mental model of real or imagined functions; III the model adapts through contextualized interaction (e.g., at home).
- **Integrative framework of anthropomorphism (IFA)** (Spatola et al., 2022): includes individual and **cultural** variables, e.g., Japanese **animism** (things have a spiritual essence) may extend to robots.

8.3 Consequences of anthropomorphism (8.4)

- **Inflated expectations:** humanlike appearance leads people to expect gesture, gaze, meaningful interaction and walking; real capabilities usually disappoint. Example: the **Geminoid** android raises high expectations, but its tele-operated reality disappoints.
- **Job suitability:** mind attribution affects perceived suitability of robots for jobs and thus deployment (Wiese et al., 2022).
- **Perceived threat:** feeling threatened in one's sense of humanness; fear of being replaced, outperformed or monitored (privacy and data misuse). Conflicting evaluations cause **ambivalence**.
- **Positive side:** humanlike perception can increase **trust** in AI, personal assistants, autonomous vehicles and HRI.

Trust in technologies (8.4.1)

- Trust = confidence that a person or system will act appropriately. **Sabel (1993):** *the mutual confidence that no party to an exchange will exploit another's vulnerabilities.* Implies trust in benevolence, integrity, trustworthiness and performance.
- **Automation** = a machine carrying out a function previously done by a human (Parasuraman & Riley, 1997); trust research emphasised system performance.
- HRI trust (Hancock et al., 2011, 2021; Kessler et al., 2017): determined by **robot, human and environmental factors**; meta-analysis stresses **human-related factors**. No single agreed definition, but many scales exist.

Accepting robots (8.4.2)

- **TAM (Technology Acceptance Model):** willingness to use depends on **perceived usefulness** and **perceived ease of use** → intention to use → use. Treats the robot as a tool; ignores context (de Graaf et al., 2019).
- **UTAUT:** ease of use → **effort expectancy**; usefulness → **performance expectancy**; adds **social influence** and **facilitating conditions** (environment). Used for child–robot learning.
- **Almere model** (Heerink et al., 2010): adds affective factors (**trust, perceived enjoyment, attitude**); developed for seniors and assistive technology.
- **User experience (UX):** adds hedonic factors (e.g., stimulation), social norms and control beliefs; proposes technology qualities that shape perceptions. TAM + UX merging has been proposed.

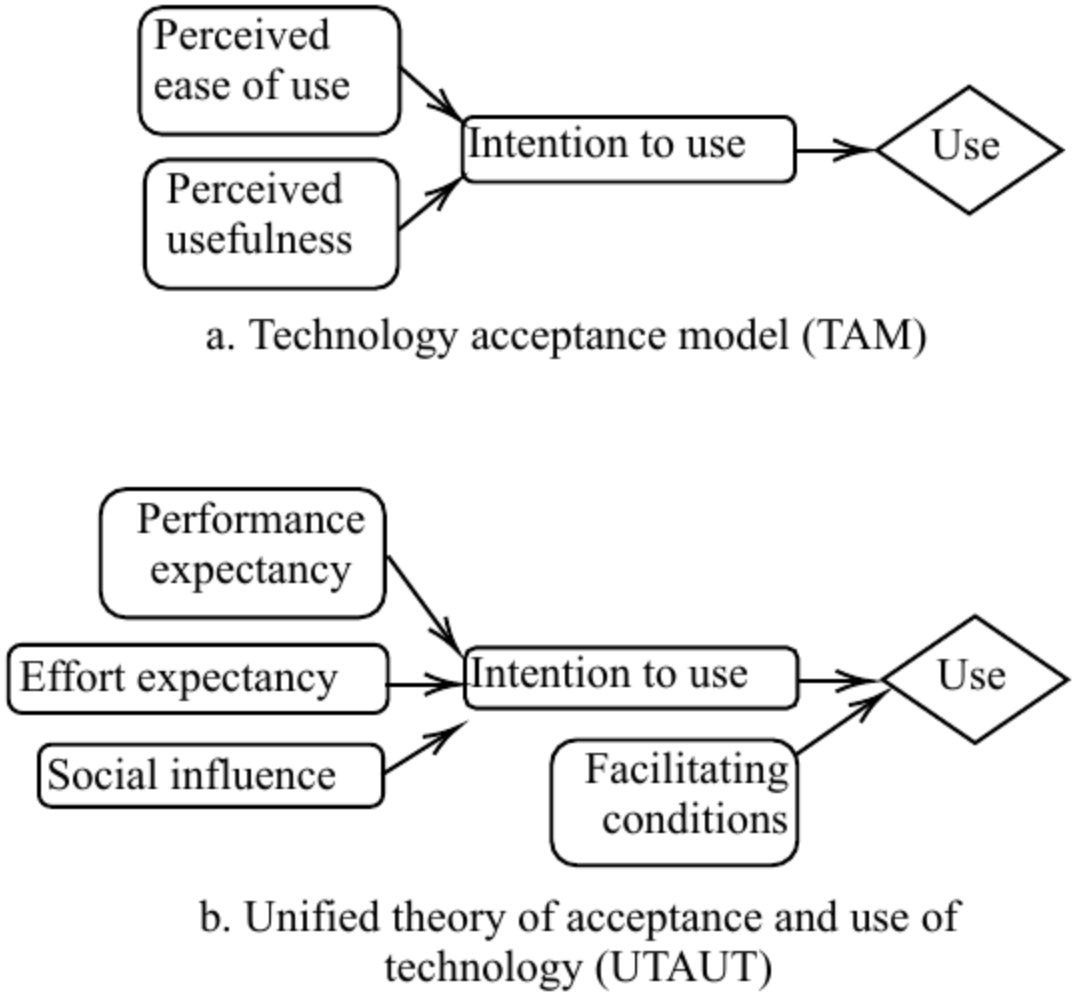


Figure 8.3 (textbook): (a) Technology acceptance model (TAM): perceived ease of use and perceived usefulness lead to intention to use, then use. (b) UTAUT: performance expectancy, effort expectancy, social influence and facilitating conditions lead to intention to use and use.

Liking and disliking robots (8.4.3)

- **Likability** = affective evaluation of how pleasant or appealing a robot is; linked to willingness to collaborate, being persuaded and prosocial behaviour.
- **Uncanny valley** (Mori, 1970): likability rises with human-likeness, then drops when almost-but-not-quite human.
- Recent view: the drop may come from a **mismatch** between features (e.g., humanlike skin with slightly-off facial movement; Kätsyri et al., 2015), also seen in zoomorphic robots and mixed gender cues. A **novelty factor** exists: uncanniness fades with short- and long-term interaction.
- Likability increases with anthropomorphism, emotional cues, and **movement in sync** with the user; more natural robot "skin" is liked more (Yamashita et al., 2016).
- Takeaway: perception of a robot and its actual appearance/function **interact**.

9. Quick Revision Points

- Nonverbal types: **gaze, gesture, mimicry/imitation, touch, posture/movement, rhythm/timing**.
- Gesture classes: **deictic, iconic, symbolic, beat**.
- Mimicry = unconscious; imitation = conscious; chameleon effect = persuasion by mimicry.
- Midas effect: touch → bigger tips.
- Gesture generation: **key frames, motion capture, BML, ML/deep learning, cognitive architectures**.
- Speech units: **utterance, word, phoneme**; conversational fillers keep flow.
- ASR: time domain → frequency domain → deep-learning transcription; problems are noise, distance, accents, proper nouns; fixes include close miking, constraining vocabulary, ROILA, VAD.
- Turn-taking: back-channeling + timing; human responses ~100–500 ms, robots take seconds; use fillers, just-in-time synthesis, incremental processing.
- Dual-process: automatic (System 1) vs deliberate (System 2).
- Epley's three factors: **effectance, sociality, elicited agent knowledge**.
- Acceptance models: **TAM → UTAUT → Almere → UX**.
- Affect = umbrella term; emotion = specific cause, short-lived; mood = diffuse, longer-lasting.
- Emotion models: Ekman six expressions, OCC (22 categories), Russell 2D (arousal, valence), PAD 3D (pleasure, arousal, dominance).
- Robot emotion strategy: mimicry (except anger); expectation management matters.
- Perceiving emotion: face (computer vision), gait, speech prosody, skin conductance. Expressing: FACS action units (AU6 + AU12 = smile), body, motion, light, sound.
- Affective HRI challenges: context, acted training data, subtle emotions, limited categories, individual differences, no long-term models, robot does not truly feel.

10. Likely Exam Questions

1. Explain the functions of nonverbal cues in HRI with examples.
2. Describe the different types of nonverbal interaction and how each is used in robots.
3. How do robots perceive and generate nonverbal cues? Explain key-frame animation and cognitive approaches.
4. Differentiate mimicry and imitation; explain the chameleon effect and its use in robots.
5. Explain the principles of speech recognition and its limitations in HRI. What are VAD and ROILA?
6. Discuss turn-taking in HRI: back-channeling, timing and solutions to response delay.
7. Explain the dual-process model of impression formation.
8. Describe Epley's three determinants of anthropomorphism.
9. Compare TAM and UTAUT. What are the Almere model and UX additions?
10. Discuss the consequences of anthropomorphism: expectations, threat, trust, likability and the uncanny valley.
11. Differentiate affect, emotion and mood. Why are emotions important in social interaction?
12. Explain primary and secondary emotions and Russell's circumplex model with a diagram.
13. Describe the OCC and PAD emotion models. Why is PAD preferred over the 2D circumplex?
14. How do robots perceive human emotions (face, speech, other modalities) and how do they express emotion (FACS, non-anthropomorphic cues)?
15. What are the challenges in affective HRI?