Chapter 15
Gestures in Educational Behavior Coordination. Grounding an Enactive Robot-Assisted Approach to Didactics

Hagen Lehmann and Pier Giuseppe Rossi

Abstract The ability to coordinate behaviors at an interindividual level has shaped human social evolution by enabling the formation and maintaining the cohesion of large social groups. Nonverbal communication always played a central role in this process—a role that now can be expanded. With the introduction of social robots, capable of emulating human appearance and movements to communicate with us through social signals, the mechanisms of human–human nonverbal communication offer us a way to improve human social communication with robots in a variety of fields—from information to assistance to people with special needs. In this article, we explore this possibility with reference to educational robotics, and, more precisely, to robot-supported didactics. In the first part of this chapter, we discuss the concepts of behavior coordination and structural coupling as evolutionary mechanism underlying human social structures and illustrate the importance of nonverbal communication in social interactions. We will give examples of different nonverbal communication channels and illustrate, with recent paradigmatic studies, how they can be used for social robotics in different cultural settings. In the second part of the chapter, focusing on educational robotics, we illustrate how nonverbal communication between humans and robots can be used as feedback channel between teachers and students in order to reinforce the structural coupling in an enactive robot-assisted approach in didactics.

15.1 Behavior Coordination in Human Evolution

Human social evolution is, to a large extend, driven by the human capability to communicate about past experience, and in this way to pass on and to accumulate cultural techniques [1, 2]. Humans transmit information to each other via a plethora of different signals. These signals can roughly be categorized into verbal and nonverbal. Verbal signals include language and utterances, like shouts and laughter. Nonverbal signals include touch, facial expressions, body posture, and gestures.

H. Lehmann (✉) · P. G. Rossi
Università degli Studi di Macerata, Macerata, Italy
e-mail: hagen.lehmann@unimc.it

While communicating, humans exhibit a multiplicity of these nonverbal behaviors at the same time, and many of them are displayed subconsciously. The expression of these behaviors, as well as their recognition, involves almost the entire body [3]. Humans are able to use the posture of conspecifics, the way they move in terms of speed and expressivity, their tone of voice and general appearance to deduce or even understand internal states like emotions or level of arousal. This understanding enables us to feel empathy for one another [4], which plays an important role in the formation and maintenance of social cohesion in large groups of individuals [5], like human societies. Since most of the cues used to "understand and feel for" the other are nonverbal, the importance of nonverbal communication for human social evolution cannot be overestimated [6]. Face, eyes, and hands play a central role in this process [7]. Crucial for the interaction with others are subconscious eye movements like gaze and pupil dilation and hand and arm gestures [8]. Most of these nonverbal signals have facilitating, regulating, and illustrating functions [9], and are as such part of the embodied information exchange that makes coordinated communication between two or more people possible.

15.1.1 Embodiment and Structural Coupling

Humans can be represented as complex self-organizing systems dynamically embedded in complex self-organizing environment(s) [10]. In this theoretical perspective, the process of adaptation is often thematized in terms of "coevolution." The general idea is that of an dense interaction, made of exchanges of energy and matter, between two operatively independent self-organizing systems. Typically coevolution is characterized as a symmetrical relation of reciprocal perturbations and endogenous processes of self-regulation that coordinates the dynamics of a system with the dynamics of its environment. Until both these two systems maintain their organization, the dynamical evolution of each of them consists of a series of endogenously generated states of activity that are compatible with the self-organizing states of the other system. Humberto Maturana and Francisco Varela, within the theory of autopoiesis, offered a particularly well-defined notion of coevolution in terms of "structural coupling" [11]. Introduced by Maturana and Varela to conceptualize the adaptive coupling as a cognitive coupling, this notion indicates the capability, typical of biological systems, to effectively act within their domain of existence to maintain and develop their organization and their mode of existence. According to the theory of autopoiesis, at the level of the dense interactions between conspecifics characterizing social environments, structural coupling becomes "behavioral coupling": a symmetrical relation of reciprocal perturbation and endogenous self-regulations that generates the interdependence of the behavioral conducts of the interacting systems. In humans, behavioral coupling is the basic structure of social interaction based on communication [11].

When developing enaction in the 1990s (e.g., [12]), Varela put this notion of structural coupling at the center of his theory. The loop shown in Fig. 15.1 illustrates
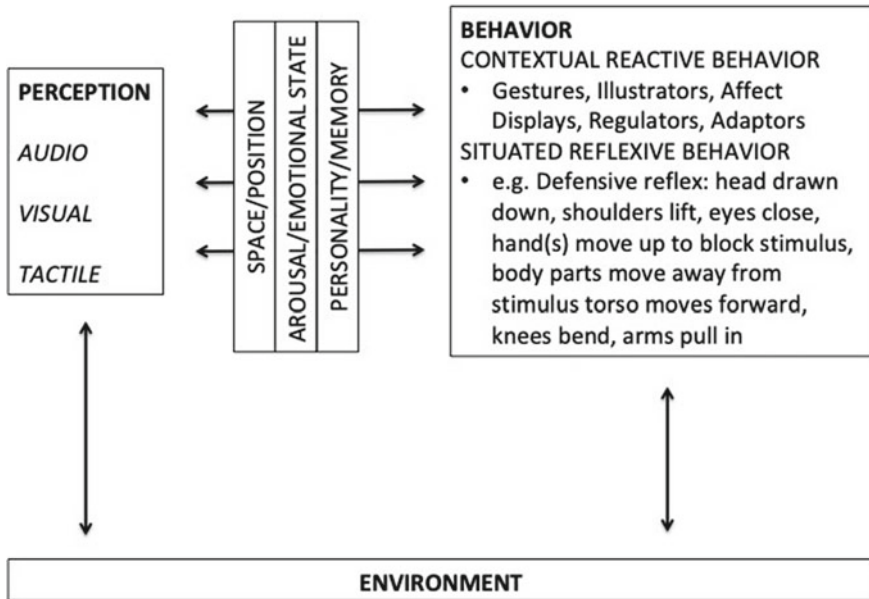
Fig. 15.1 Structural coupling between environment and human perception and behavior

the structural coupling between an individual and its environment. The changes in the dynamics of the environment generate perturbations in the dynamic of the system, which reacts on these changes via different self-regulative behaviors to compensate them. These behaviors generate in turn perturbations in the environment, and so on. In case of social interactions between two or more humans, the internal equilibria can be represented also by the individuals personality, which depends on the individuals phylo- and ontogenetic history, and the perceptible changes can be represented by the different verbal and nonverbal communication signals.

In order for a social exchange to be successful, i.e., to achieve a common goal, which in its simplest form could mean to have a conversation, the behaviors of the individual and its environment need to be coordinated [13]. This type of coordination can be found on all levels of embodied behavior, from eye movements [14] to coordinated neuronal patterns [15].

In order for robots to be accepted into mixed human–robot ecologies [16], it is important that not only their verbal, but also their nonverbal behavior is aligned with the expectations of the users. As pointed out above, human nonverbal behavior incorporates a multitude of signals. Specifically for robots that are operating in close physical and even social proximity to humans, the same should be true. For example, it has been shown that different robot blinking patterns can influence how the robot is perceived [17]. This is even more true for contextual reactive behaviors like gestures.

Research has shown that with an increasing level of autonomy and human likeness in appearance of robots, their human users have the tendency to anthropomorphize

them [18, 19, 20] (Damiano and Dumouchel; Eyssel and Kuchenbrandt). Since the goal of social robotics is to enable intuitive and comfortable interaction with between robots and humans, robots should be enabled to become part of the structural coupling of humans and their environment by endowing them with capabilities of behavior coordination. In other words, if we understand both human–social interactions, and human–robot interactions as coevolutionary processes, or processes of structural coupling, we can apply the principles of enaction in the design process of robotic behaviors. In the second part of this chapter, we will discuss the implication of enaction further from an educational perspective.

Research on social coordination shifted into focus of evolutionary anthropology in the middle of the 1960s. One important task was to find a categorization for nonverbal behaviors that explained many of the observed phenomena and allowed for predictions of group dynamics. Ekman and Friesen [21], for example, separated nonverbal behaviors into contextual reactive and situated reflexive.

15.1.2 Reflexes

According to Ekman and Friesen's definition, the latter included the orientation reaction and the startle reflex, if something or someone touches us or appears quickly and unexpectedly in the personal zone of a person [22]. In this case, the person unwillingly draws the head in and lifts the shoulders to protect the neck, closes the eyes to protect them, draws the arms in and moves the hands up to protect the body, bends the knees slightly, and moves the body away from the stimulus [23]. Another reflex in this category would be the orientation reaction, which is exhibited when an unexpected event occurs around a person not fast enough to initiate the startle reflex. In this case, the person's body will stiffen, and the person will orient herself toward the stimulus and exhibit a general outward alertness [24]. On the other hand, there are reactive contextual behaviors, which are usually used to influence conversation dynamics. They can have an illustrative function emphasizing what is currently said, a regulatory function facilitating turn-taking during conversations, or they can specific linguistic meaning like most hand gestures.

15.1.3 Facial Cues

For humans, the highest concentration of different sensors is located in the face, harboring the mouth, the nose, the eyes and to a certain extent the ears as sensory input channels is also the focal point when communicating with conspecifics. As highly visual species humans automatically "face" their counterpart when they want to start a social exchange or when they are addressed by someone else, in order to see her/his intentions. Since the hairless human face allows for the visibility of very small muscle movements, it is not surprising that facial expressions are one of the most

efficient channels for the transmission of information about the emotional states of the other, and that a lack of facial expressivity creates in humans a sense of eeriness. Social eye movements like gaze following, change of pupil size, and blinking have been shown to be among the most powerful signals humans use to create, maintain, or disturb group cohesion or peer-to-peer interaction [25, 26]. The specific visibility of the human eye [27] turns it to a communication channel that is unique in nature.

15.1.4   Gestures

Despite the importance of the above-mentioned communication channels, the importance of the hand and arm gestures for nonverbal communication is central. When engaged in social exchanges, in which one is not required to have "ones hands full," the hands are usually used to illustrate and emphasize what is currently said and even thought, as well as to regulate the conversational dynamics of an interaction. This is usually done via a set of cultural depending gestures. These gestures are essential for ensuring comfortable and intuitive social exchanges. In contrast to other subconscious nonverbal communication signals, gestures are population dependent [28, 29, 30].

Communicative gestures have evolved in different parts of the world, which were isolated from each other for long periods of time. This, in combination with the physical constraints of the human body, led to the effect that the same gesture can have very different meanings in different cultures. However, it is important to point out that despite these differences it is possible, albeit on a very basic level, to establish communication via gestures between members of very different cultural backgrounds. This hints at the long evolutionary history and importance of gestures as communication channel in human evolution. In some cases, the differences can be quite striking. For example, going from Europe to Japan and seeing a Japanese person waving her hand in front of her face with the face turned toward you could lead to quite a severe misunderstanding. This gesture, in Europe commonly understood as an insult with the meaning "Are you crazy?" is meant as an apologetic negation in Japan (Fig. 15.2).

But even within Europe, the differences are very noticeable. In southern Europe, namely in Italy, gestures are used much more frequently during conversations when compared to countries of northern Europe. Comparing the frequency and expressivity in the use of hand gestures during a discussion among Scandinavians or among Italians would illustrate the point (Fig. 15.3).

These examples show that gestures, which have played a crucial role during the early social evolution of our species, remain very much alive in human social communication. Research exploring different aspects of human cognition has demonstrated the universal importance of gestures for enhanced information transfer [32] and lexical retrieval [33]. It has even been shown that using gestures helps to reduce the cognitive load when explaining complex problems to others [34]. In this way, gestures not only reflect our cognitive state, but also shape it.

Fig. 15.2 Examples of Japanese communicative gestures (from [31]). Starting from top moving clockwise the gestures mean no (waving hand in front of face), I (pointing to nose), money, and apology for intruding personal space of other

Fig. 15.3 Examples of Italian communicative gestures (from [31]). Starting from top moving clockwise the gestures mean What is going on? something tastes very good, moderate threat, aggressive disinterest

One of the theories about the origins of human language is the gestural origin hypothesis [35]. It proposes that the use of gestures predates the evolution of verbal language. There is archeological, physiological, and behavioral evidence that support this theory. For example, paleo-archeological findings show differential growth in the brain and the vocal apparatuses [36]. Human babies exhibit gestural communication before they speak [37]. Bonobos and chimpanzees use gestures to communicate nonverbally without touching one another [38]. Apes and humans show a bias toward the usage of the right hand (left brain) when gesturing [39, 40]. In apes, the Brodmann area 44, a brain region that is activated during the production and perception of gestures, is enlarged in the left brain hemisphere [41].

These findings illustrate the high relevance of gestures for human–human communication. Gestures are deeply rooted in primate social evolution. In combination with facial expressions and vocal signals typical of apes and humans, they added a layer of flexibility to the behavioral repertoire that allows for great communicative complexity, which drove human social evolution.

15.1.5   Gestures in Human–Robot Interaction

The understanding of the importance of nonverbal communication, in combination with the technological progress of robot embodiments that allow the expressions of nonverbal signals, has lead in recent years to various approaches to implement and test communicative gestures in humanoid and non-humanoid robots. These implementations were done from different perspectives and were based on different research questions. In this section, we will discuss exemplary studies that aimed at developing gestures and other forms of nonverbal communication for different robotic environments.

Ono et al. [42] presented in their work a model of embodied communication, including both gestures and utterances. They tested their model with the Robovie platform, in an experimental setup in which the robot gestured to various degrees while explaining the route to a designated goal to a human interlocutor. They could show that (a) the more the robot gestured systematically, the more the human subjects' gestures increased in frequency, and (b) that the more the robot used gestures, the more the better the humans understood its utterance about how to reach the goal. Other research examined the role of gestures in the process of starting an interaction with a robot, maintaining it, and perceiving a connection to one another [43]. The results of these experiments showed that people direct their attention more frequently to robots and find their interactions with the robot more appropriate when gestures are present in the interaction. Riek et al. [44] tested the effect of different aspects of interactional gestures made by a robot on the ability of humans to cooperate with this robot. They found that humans were cooperating quicker when the robot made abrupt, front-oriented gestures.

Beck et al. [45] tested whether it is possible with for a robot to express emotions with body language in such a way that children are able to understand and interpret

them. They used different body postures of the robot for typical emotional states like happiness, fear, anger, and pride. Their results underlined the importance of the position of specific body parts, i.e., the head position, during the expressed emotion in order to ensure the interpretability of the expression.

Another very interesting insight into how to use the body language and gestures during human–robot interaction comes from [46]. They use different gestures and gaze behaviors in order to test the persuasiveness of a storytelling robot. In their experiment, the participants listened to a robot telling a classical Greek fable. Their results showed that only a combination of appropriate social gaze and accompanying gestures increased the persuasiveness of the robot. The authors pointed out that in the condition the robot was not looking at the participants and only used gestures, the persuasiveness of the robot actually decreased because the participants did not feel like they were addressed.

This illustrates an important point for future HRI research. It is not sufficient to look only at different aspects of body language and then to model them separately on the robot, but it is at least as important to focus on their integration in order to achieve a holistic behavior expression during the interaction. Using video footage of professional actors, as was done in this study, is a good starting point for the modeling of these dynamics. Huang and Mutlu [47] used a robot narrator equipped with the ability to express different types of gestures. They designed deictic, beat, iconic, and metaphoric gestures following McNeill's terminology [32]. The results showed interesting effects for the different types of gestures. Deictic gestures, for example, improved the information recall rate of the participants, beat gestures contributed positively to the perceived effectiveness of the robots gestures, and iconic gestures increased the male participants' impression of the robot's competence and naturalness of the robot. An interesting aspect of their findings is that metaphoric gestures had a negative impact on the engagement of the participants with the robot. The authors state that a large number of arm movements involved in this type of gesture might have been a distraction for the participants.

These studies illustrate that researchers in HRI have recognized the importance of gestures for their field. Besides the insights this research gives into how humans use and understand gestures, and it also has a very practical and applied use. Specifically, the last five years have seen the deployment of a multitude of social robotic platforms in areas that range from shopping malls to schools and airports [48]. International projects like the Mummer project [49], for example, experiment with social signal processing, high-level action selection, and human-aware robot navigation by introducing the Pepper robot in a large public shopping for a long-term study. The result of this project was applications that enable the robot to talk to and to entertain customers with quizzes, and give guidance advice by describing and pointing out routes to specific goals in the shopping mall.

These examples illustrate that social robot need, for almost all of their future applications, to be able to interact with humans in human terms. Once the robots have left the laboratory and the factory, their communication capability needs to be appropriate for laymen users, i.e., they need to make themselves understood in an easy and intuitive way.

As pointed out on page 4, the frequency and type in the use of gestures are culturally dependent. If we imagine a social robot that is, for example, built in Europe, equipped with gesture libraries based on northern European social interaction dynamics and sold worldwide, it is easy to understand the issues that could arise. It is therefore important to stress that it is necessary to not only understand how to design gestures for social robots, but also to conduct comparative research and develop cultural sensitive gesture libraries. The result of an earlier study that was aimed at establishing a baseline for robot gestures during human–robot conversations [50] demonstrates this need. During the study, conversational pairs of humans were videotaped and their use of gestures was analyzed and compared. The research was conducted in Italy and in Japan, respectively. In this research, gestures were defined as nonlocomotory movements of the forearm, hand, wrist, or fingers with communicative value, following definitions from other behavioral research [38, 51], and communicative movements of the head like nodding up and down, shaking left to right, and swaying. The results showed expectedly quite severe differences not in the type, but also in the frequency and expressivity of the gestures used. Italians used their arms and hands considerably more during the conversations than the Japanese participants. While Italians used much more iconic and metaphoric gestures, the Japanese participants used small head movements to control and regulated the conversational dynamics.

Other studies found similar effects between participants from different cultural backgrounds.

Trovato et al. [52], for example, researched the importance of greeting gestures in human–robot interaction between Egyptian and Japanese participants. They could show that specifically during the robot's first interaction with a human it can be crucial to have a culturally sensitive gesture selection mechanism. They argue that once social robots will become mass-produced products, its cultural sensitivity in the behavior of the robot will determine its success rate. If users have the possibility to choose the robotic platform they are most comfortable with, then it stands to reason that they will choose one that exhibits cultural closeness. In another study, the same group presented a cultural sensitive greeting selection system [53]. Their system was able to learn new greeting behaviors based on their previous Japanese model. The research was conducted with German participants and the results showed that the model was able to evolve and to learn movements specific to German social interaction dynamics. The authors argue that this type of cultural sensitive customization will become more and more important and that robots should be able in the future to switch easily between different behavioral patterns depending on the cultural background of the human user.

In this first part of the chapter, we illustrated the importance of nonverbal communication and behavior coordination in human–human communication from a social, anthropological, and evolutionary perspective and showed how gestures, as one type of nonverbal-social signal, can be used during human–robot interactions. This is the framework in which we contextualize the second part of the chapter, which discusses an implementation of the theoretical concepts of behavior coordination and enaction in educational robotics.

15.2 Robots in Education

The previous part of this chapter was intended to give an overview of the role nonverbal communication and behavior coordination played in human social evolution, and to illustrate why the use of nonverbal communication signals for social robots that need to interact with humans in close physical and social proximity is important for the success of this technology. We looked at human–robot interaction research and saw an increasing awareness of the importance of social gestures for the field. In the following part, we will look at one field educational robotics and explore how social robots can be implemented in the teaching process and what role nonverbal communication and behavior coordination can play for the success of these robots. We will propose a new didactic framework, which represents an extension of the enactive approach to didactics [54] and ascribes to social robots a central role in the feedback process between teachers and students. It will become clear, why the use of robotic gestures in this framework is essential for the success of the enactive approach.

15.2.1 From Tools to Mediators

Since the development of Lego Mindstorms NXT [55], an increasing number of robots, have been deployed in schools, not only to teach programming, but also scientific subjects like physics or chemistry (e.g., [56, 57]). The integration of the Lego Mindstorms into school curricula followed a "constructionist" framework and the related "learning-by-making" methodology, as it was originally proposed by [58]. It has mainly been used in middle schools and high schools to teach students the basic principles of what robots are, how they work, and how software applications can be developed for them [59, 60]. This kind of uses of robot technology in schools enforced the kind project-based learning strategies [61], in which teachers usually engage their students into artifact or product building activities, and which we still see most frequently in technology-assisted STEM education.

However, the last ten years have seen more and more social robots being integrated into, for example, primary school language classes and in robot-assisted therapy settings for children with special needs. These robots are usually humanoid and serve in the function of social mediator.

As pointed out in section "Embodiment and Structural Coupling," in order for social interactions to be successful, behavior coordination is central. This is specifically true in educational contexts. Hence, mechanisms to provide appropriate feedback from robots in tutoring situations have moved into the focus of research on social robots in education (e.g., [62]). This feedback is usually based on different sensory inputs from human social signals, and on the processing of these social signals. Social signal processing with the goal of improving robot feedback has been at the center of various recent social robotic projects [49, 63].

In the specific case of long-term interactions between robots and children, the issue arises that the novelty effect of using robots wears off quickly and that the children subsequently become bored. In these circumstances, the robot does not only need to be reactive in a specific task, but additionally, it needs to provide appropriate emotional feedback. This kind of feedback needs to be based on memory models of the children's behavior over time. First successful attempts in this direction have been made to support vocabular learning in primary school students [64].

Different ways of classifying robots in educational contexts have been. For example, Mubin et al. [65] and Tanaka et al. [66] identify two different ways in which robots have been integrated into school curricula. As pointed out above one is as educational tools in themselves, e.g., to teach children the basic principles of programming, and one as educational agents. The latter category includes social robots like, for example, RoboVie [67], Tiro [68] and NAO [69]. A further classification of the roles of social robots in educational contexts has recently been given by Belpaeme et al. [70]. In their review, they found that this kind of robots mainly fulfills the roles of novices, tutors, or peers. When fulfilling the role of novice, a robot allows the students to act as tutor and to teach the robot a determined topic. This helps the children to rehearse specific aspects of the syllabus and to gain confidence in their knowledge [71, 72]. When the robot is fulfilling the role of tutor, its function is usually that of assistant for the teacher. Similar to robotic novices, robotic tutors have been used in language learning classes. Strategies used in robot-based tutoring scenarios include, for example, encouraging comments, scaffolding, intentional errors, and general provision of help [73]. The idea behind having robots assume a peer role for children is that this would be less intimidating. In these cases, the robot is presented as a more knowledgeable peer that guides the children along a learning trajectory [70], or as an equal peer that needs the support and help of the children [71].

Another very important field in which robots have been used to achieve educational goals is robot-assisted therapy (RAT) for children with special needs. Robots like KASPAR [74] fulfill the role of social mediator to facilitate social interaction among and between children with autism spectrum condition (ASC) (e.g., [75]). In this function, the robot teaches the children appropriate social behaviors via appropriate verbal and nonverbal feedback. RoboVie R3, on the other hand, has been used very successfully in the teaching of sign language to children with hearing disabilities. For this purpose, it was equipped with fully actuated five-fingered hands. In their study, from 2014, Köse et al. [76] describe comparative research between NAO and RoboVie R3. The mode of interaction between the robots and the participants was nonverbal, gesture-based turn-taking, and imitation games. Their results showed that the participants had no difficulty to learn from the robots, but that they found it easier to understand Robovie R3's performances due to it having five fingers, longer limbs, and being taller than NAO. These findings could be seen as evidence that for gesture-based communication, child-sized robots like RoboVie R3 and Pepper might be in an advantage given their better visibility and the apparent better interpretability of their movements. In follow up studies to their original research, Köse et al. [77] and Uluer et al. [78] replicated their original results using RoboVie R3 as an assistive

social companion in sign language learning scenarios. They could additionally show that the interaction with the physical robot is more beneficial for the recognition rate of the gestures performed by the robots, when compared to a video representation.

As shown in Fig. 15.4, social robots are used in an area in which they are not considered as tools, i.e., subjects and part of the knowledge to be transmitted, but in the area where they are directly or indirectly transmitting knowledge. The function of the robot changes from object to educational agent involved in the generation of new knowledge. This moves the robot into the center of the teaching process. As we discussed on page 1 of this chapter, human culture has a cumulative nature and our social evolution is "ratcheted up" by active teaching [1]. This process is inherently human and the cultural techniques linked it to follow a trajectory that intuitively connects individuals and increases social cohesion in groups. They are necessarily based on verbal and nonverbal communication techniques and involve the entire human repertoire of social signaling. If we ascribe robots an active function in this process, it stands to reason that they need to be equipped at least to some extent with the capability to use body language and gestures.

Following this line of thought, it is noticeable that a lot of robots that are used as educational agents are either humanoid or semi-humanoid, such as NAO, Robovie R3 [79], or Maggie [80]. One of the reasons for this is that human features like a moveable head, moveable arms, and actuated hands are most suitable for the implementation of human nonverbal communication signals. However, this makes the development and implementation of this kind of fully embodied agents in education much more costly and difficult, than the use of robots similar to the ones that can be constructed from Lego Mindstorms. Herein lies the reason why, until now, the majority of robotic
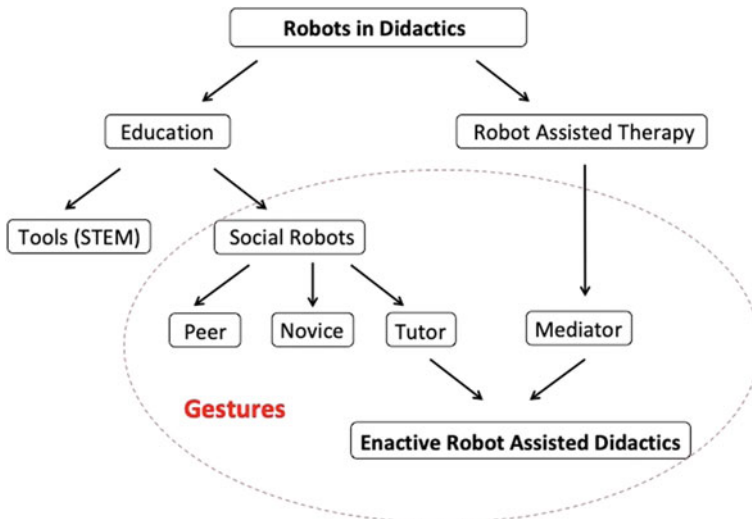

Fig. 15.4 Roles of robots in didactics. The red oval marks the space in which we propose robots should use gestures

technology was used as tools for STEM education in the past [81, 82]. However, with the readily availability of robots like NAO or Pepper, this is changing. These new types of robots lent themselves to be integrated into new existing theoretical approaches in the field of didactics. On such approach that gains momentum at the moment is enactive didactics. A detailed description of the enactive didactics approach can be found in Lehmann and Rossi [83].

15.2.2   *Enactive Robot-Assisted Didactics*

The enactive didactics approach focuses on the interactions between teacher and student during the knowledge creation process. The teacher is seen as the focal point that raises the awareness of an issue in the students. In the next step, the teacher and the students build an answer to the issue together. The trajectory along which this answer is constructed and sketched out by the teacher. She has the role of mediator between the world of the student and the new knowledge [84], and the task of activating a cognitive conflict [85] that bridges the student's knowledge, the new problems to address, and related new knowledge. After the new knowledge is established, it is crucial to validate it. In the enactive didactics approach, it is the function of the teacher to verify the epistemological correctness of the constructed knowledge, ensuring that it does not contradict the existing knowledge. In order to establish this validation, continuous feedback between the teacher and the students is necessary. The role of feedback is not only important for the student in this process, but also for the teacher, as each part of the teacher–learner dyad is seen as part of the structural coupling between the environment and, respectively, the teacher and the students (see Fig. 15.5a). Unfortunately, in reality, many interaction processes in education lack the space for interaction and feedback for various reasons. This absence of real feedback, however, produces self-referentiality, which is a characteristic of closed systems and diametrically opposed to the form of interaction between a subject and its environment as it is described in the enactive approach.
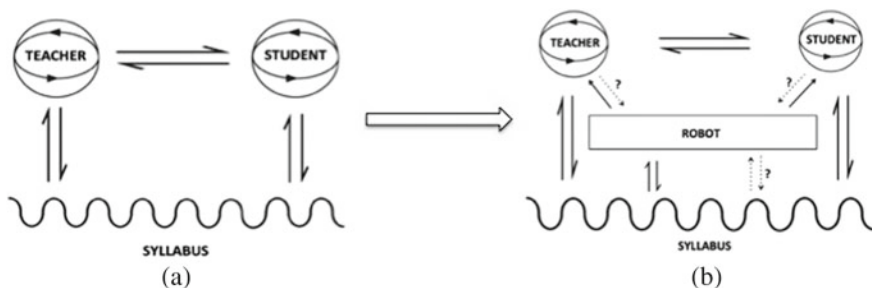
Fig. 15.5 Extension of the structural coupling characterizing the enactive didactics approach by integrating a robotic tutor (taken from Lehmann and Rossi [86])

As we proposed analytically elsewhere [86], integrating social robotics technology based on the enactive framework has the potential to remedy the problematic lack of feedback by reinforcing the reticular interactional structure described in the approach (see Fig. 15.5b). In other words, the integration of a robot in the function of social mediator will strengthen the communication between teacher, students, and syllabus (knowledge to be thought). Consequently, we describe this approach as enactive robot-assisted didactics (ERAD).

The central point of this idea is the strengthening of the communication between the human actors. In order for the robot to be successful, its attempts to initialize communication have to be intuitively understandable and most importantly nonintrusive or disruptive. The robot must be capable to catch the attention of the teacher or the students without disturbing the flow of the lecture and to intervene in a way that is perceived constructive and helpful.

In order to achieve this, we need to shift our attention to human–human nonverbal communication. As discussed before in this chapter, humans have an entire evolutionary history of using body posture, and more specifically head, arm, and hand movements to seek attention and transmit information to conspecifics. If robots ought to be successful in social mediator functions like the ones described here, they need to be enabled to tap into this behavioral repertoireand exploit the evolved human abilities to interpret the body movements of other. Since this ability to "read" our counterpart is limited, other humans,[1] this type of robots should be either humanoid or semi-humanoid (i.e., they should have a head, arms, and hands).

For ERAD, we propose a number of techniques that will enable the robot to collect data from the student and the teacher, but the central part is the communication abilities of the robot. Specifically, in noisy environments like the classroom, these abilities strongly depend on the robot's capability to use gestures. Since robots are already in the process of being integrated in such different cultural context like Japan and Western Europe, it will not be enough to equip robots only with one similar set of gestures. As pointed out by Trovato et al. [53], the only robots that are capable of adjusting their behaviors to a specific cultural background will be successful in an increasingly competitive market of robotic social mediators.

---

[1]Exceptions are species like dogs, with which we share a long evolutionary history and which have been bred selectively to understand human body language and to be understandable.

15.3   Conclusive Remarks

Since nonverbal communication signals and behavior coordination are from an evolutionary perspective such as important and integral part of human social interaction, it seems natural to use these concepts also in interactions with social robotic technology. It might even be necessary to rethink our approach to designing this type of interactive technology, following more a communication and coordination-driven perspective on the embodiments we construct. The research and theory discussed in this chapter underline the importance of cultural sensitive gesturing for social robots. In order for these robots to appear authentic and trustworthy, and to be intuitive to interact with, it will be necessary to equip them with a repertoire of nonverbal communication behaviors that are adequate for the cultural context they are used in. We argue that the way forward is a detailed analysis of the cultural specificities of each general population in order to generate the necessary behavioral libraries. Behavioral anthropologists have, for example, listed and described many cultural-specific gestures (e.g., [31]). The results of this research could be used and implemented in social robots. However, it is not enough to equip robots with specific executable, but their motion dynamics and frequencies in dependence of the reactions of their recipient need to be taken into consideration.

   We chose the field of educational robotics for the illustration of how social robots could assume a central role in human interaction dynamics. The examples from educational robotics show the possibilities social robotic mediators and tutors have to ease and facilitate the approaching didactic shift caused by the rapid technologization of learning environments. Specifically, Asian countries like Japan, South Korea, and Singapore have embraced the use of robots in pre-schools and schools. Robots like TIRO and Robovie have been integrated in the school curricula and are supporting teachers in the classroom. The majority of the applications of these robots are linked to language learning and involve the robots linking new words and grammatical concepts to movements and gesturing and in this way multimodal anchoring the new knowledge in the memory of the children.

   In order to put these applications on a sound theoretical didactic basis, we propose an extension of the current enactive didactics approach. We suggest to ascribe to social robots a central role in the feedback process between teacher and students in order to reinforce the reticular character of the structural coupling during the learning process. We argue that this central role requires from the robots embodied nonverbal communication competencies, whose character should be similar to this of humans to be easily understood and nondisruptive. This need for human similarity to human means that robots should be equipped with culturally sensitive social gesture libraries, which can be expressed best with a humanoid or semi-humanoid embodiment. A convergence in this point would also bear a further advantage. Even though there might be differences between the used robot embodiments, the general humanoid structure (i.e., head, torso, arms, and hands) would make the gestures not necessarily robot specific, but a general motion framework can be imagined, which could be

used across platforms, similar to the Master Motor Map framework proposed by the KTU [87].

We plan to implement these ideas in a first step with the Pepper robot from Softbank Robotics. In order to develop and expand our enactive robot-assisted didactics approach, we are using Pepper with two main functionalities: (a) to give feedback about the structure of an ongoing lesson and (b) to enforce feedback between the teacher and students.

In scenario (a) Pepper helps, on one side, the teacher to maintain the predefined structure of a lecture and, on the other side, the students to understand the overall educational goal of the lesson. In order to do so, the robot gives an overview of what the content of the lecture will be at its beginning, and at the end of the lecture, it gives a summary of what has been discussed. Pepper uses gestures to illustrate the content of what it is saying. These gestures are specifically designed for the content of the lecture. During the lecture, the robot is used as an embodied timer. After a certain time, it will start to yawn. If the teacher does not react, it will move into a position that makes it appear tired. If the teacher still does not react, it will start to raise its arm, wave, and make the teacher verbally aware that it would be beneficial for the lecture to have a small break.

In scenario (b), we are using Pepper in combination with an audience response system (ARS). The ARSs are used for direct real-time feedback. Although their usefulness is undeniable, the feedback they provide, in form of simple statistics, is inherently unembodied and depends strongly on the willingness of the presenter to let the audience interfere with the presentation. We are using the robot in order to add an embodied component and enforce the integration of the feedback. For this concrete scenario, the lecture is structured into different sections. Each section is concerned with a specific topic. At the end of each section, the robot prompts the teacher to let the students fill in a short questionnaire about the content of the section in Google Forms with their mobile phones. After the data is collected, the robot then gives embodied feedback about the results. The prompting as well as the feedback is composed of verbalizations and informative gestures of increasing intensity.

These two examples illustrate the potential use of robots as embodied feedback devices and social mediators between students and teacher have. Many other scenarios are imaginable. The development toward a more and more embodied interaction with robots will generate intertwined human–robot ecologies, which will have potentially a profound impact on the social evolution of our species (e.g., [19]).

References

1. Tennie, C., Call, J., & Tomasello, M. (2009). *Ratcheting up the ratchet: On the evolution of cumulative culture. Philosophical Transactions of the Royal Society B, 364.*
2. Tomasello, M. (1999). *The cultural origins of human cognition.* Cambridge, MA: Harvard University Press.
3. Scheflen, A. E. (1972). *Body language and the social order; communication as behavioral control.*

4. Kacperck, L. (1997). Non-verbal communication: The importance of listening. *British Journal of Nursing, 6*(5), 275–279.
5. Van Vugt, M., & Kameda, T. (2012). Evolution and groups. *Group processes*, 297–332.
6. Burgoon, J. K., Guerrero, L. K., & Floyd, K. (2016). *Nonverbal communication*. Routledge.
7. Müller, C., Cienki, A., Fricke, E., Ladewig, S., McNeill, D. & Tessendorf, S., (Eds.). (2013). *Body-language-communication* (Vol. 1). Walter de Gruyter.
8. Argyle, M., & Ingham, R. (1972). Gaze, mutual gaze, and proximity. *Semiotica, 6*(1), 32–49.
9. Knapp, M. L., Hall, J. A., & Horgan, T. G. (2013). *Nonverbal communication in human interaction*. Cengage Learning.
10. Von Foerster, H. (2003). On self-organizing systems and their environments. In *Understanding Understanding* (pp. 1–19). New York, NY: Springer.
11. Maturana, H. R., & Varela, F. J. (1987). *The tree of knowledge: The biological roots of human understanding*. New Science Library/Shambhala Publications.
12. Varela, F. J., Thompson, E., & Rosch, E. (1991). *The embodied mind: Cognitive science and human experience*. MIT Press.
13. Oullier, O., De Guzman, G. C., Jantzen, K. J., Lagarde, J., & Scott Kelso, J. A. (2008). Social coordination dynamics: Measuring human bonding. *Social Neuroscience, 3*(2), 178–192.
14. Doughty, M. J. (2001). Consideration of three types of spontaneous eyeblink activity in normal humans: During reading and video display terminal use, in primary gaze, and while in conversation. *Optometry and Vision Science, 78*(10), 712–725.
15. Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience, 27,* 169–192.
16. Dumouchel, P., & Damiano, L. (2017). *Living with robots*. Harvard University Press.
17. Lehmann, H., Roncone, A., Pattacini, U., & Metta, G. (2016). Physiologically inspired blinking behavior for a humanoid robot. In *Proceedings of ICSR 2016*.
18. Riek, L. D., Rabinowitch, T. C., Chakrabarti, B., & Robinson, P. (2009). How anthropomorphism affects empathy toward robots. In *Proceedings of the 4th ACM/IEEE International Conference on Human Robot Interaction* (pp. 245–246).
19. Damiano, L., & Dumouchel, P. (2018). Anthropomorphism in human–robot co-evolution. *Frontiers in Psychology, 9,* 468.
20. Eyssel, F., & Kuchenbrandt, D. (2012). Social categorization of social robots: Anthropomorphism as a function of robot group membership. *British Journal of Social Psychology, 51*(4), 724–731.
21. Ekman, P., & Friesen, W. V. (1969). The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *Semiotica, 1*(1), 49–98.
22. Hall, E. T. (1966). *The hidden dimension* (Vol. 609). Garden City, NY: Doubleday.
23. Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1990). Emotion, attention, and the startle reflex. *Psychological Review, 97*(3), 377.
24. Lynn, R. (2013). *Attention, arousal and the orientation reaction: International series of monographs in experimental psychology* (Vol. 3). Elsevier.
25. Argyle, M. (2013). *Bodily communication*. Routledge.
26. Emery, N. J. (2000). The eyes have it: The neuroethology, function and evolution of social gaze. *Neuroscience and Biobehavioral Reviews, 24*(6), 581–604.
27. Tomasello, M., Hare, B., Lehmann, H., & Call, J. (2007). Reliance on head versus eyes in the gaze following of great apes and human infants: The cooperative eye hypothesis. *Journal of Human Evolution, 52*(3), 314–320.
28. Carstensen, L. L., Löckenhoff, C. E., Ekman, P., Campos, J. J., Davidson, R. J., & de Waal, F. B. M. (2003). *Emotions inside out: 130 years after Darwin's 'The Expression of the Emotions in Man and Animals'*.
29. Ekman, P. (1972). *Nebraska symposium on motivation, 1971* (J. Cole, Ed.) (pp. 207–283). Lincoln, NE: Univ of Nebraska Press.
30. Kendon, A. (1988). How gestures can become like words. *Cross-cultural perspectives in nonverbal communication* (Vol. 1, pp. 131–141).
31. Morris, D. (1994). *Body talk: A world guide to gestures*. London: Jonathan Cape Ltd.

32. McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. University of Chicago press.
33. Morrel-Samuels, P., & Krauss, R. M. (1992). Word familiarity predicts temporal asynchrony of hand gestures and speech. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*(3), 615.
34. Goldin-Meadow, S., Nusbaum, H., Kelly, S. D., & Wagner, S. (2001). Explaining math: Gesturing lightens the load. *Psychological Science, 12*(6), 516–522.
35. Corballis, M. C. (2002). *From hand to mouth: The origins of language*. NJ, US: Princeton.
36. Lieberman, P., Crelin, E. S., & Klatt, D. H. (1972). Phonetic ability and related anatomy of the newborn and adult human, Neanderthal man, and the chimpanzee. *American Anthropologist, 74*(3), 287–307.
37. Petitto, L. A., & Marentette, P. F. (1991). Babbling in the manual mode: Evidence for the ontogeny of language. *Science, 251*(5000), 1493–1496.
38. Pollick, A. S., & DeWaal, F. B. (2007). Ape gestures and language evolution. *Proceedings of the National Academy of Sciences, 104,* 8184–8189.
39. Annett, M. (1985). *Left, right, hand and brain: The right shift theory*. Psychology Press (UK).
40. Hopkins, W. D., & de Waal, F. B. (1995). Behavioral laterality in captive bonobos (Pan paniscus): Replication and extension. *International Journal of Primatology, 16*(2), 261–276.
41. Cantalupo, C., & Hopkins, W. D. (2001). Asymmetric Broca's area in great apes. *Nature, 414*(6863), 505.
42. Ono, T., Imai, M., & Ishiguro, H. (2001). A model of embodied communications with gestures between humans and robots. In *Proceedings of 23rd Annual Meeting of the Cognitive Science Society* (pp. 732–737). Citeseer.
43. Sidner, C. L., Lee, C., Kidd, C. D., Lesh, N., & Rich, C. (2005). Explorations in engagement for humans and robots. *Artificial Intelligence, 166,* 140–164.
44. Riek, L. D., Rabinowitch, T. C., Bremner, P., Pipe, A. G., Fraser, M., & Robinson, P. (2010). Cooperative gestures: Effective signaling for humanoid robots. In *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 61–68). IEEE.
45. Beck, A., Cañamero, L., Damiano, L., Sommavilla, G., Tesser, F., & Cosi, P. (2011, November). Children interpretation of emotional body language displayed by a robot. In *International Conference on Social Robotics* (pp. 62–70). Berlin, Heidelberg: Springer.
46. Ham, J., Bokhorst, R., Cuijpers, R., van der Pol, D., & Cabibihan, J. J. (2011, November). Making robots persuasive: The influence of combining persuasive strategies (gazing and gestures) by a storytelling robot on its persuasive power. In *International Conference on Social Robotics* (pp. 71–83). Heidelberg: Springer, Berlin.
47. Huang, C. M., & Mutlu, B. (2013, June). Modeling and evaluating narrative gestures for humanlike robots. In *Robotics: Science and Systems* (pp. 57–64).
48. Tonkin, M., Vitale, J., Herse, S., Williams, M.A., Judge, W., & Wang, X. (2018, February). Design methodology for the UX of HRI: A field study of a commercial social robot at an airport. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction* (pp. 407–415).
49. Foster, M. E., Alami, R., Gestranius, O., Lemon, O., Niemelä, M., Odobez, J. M., & Pandey, A. K. (2016, November). The MuMMER project: Engaging human-robot interaction in real-world public spaces. In *International Conference on Social Robotics* (pp. 753–763). Cham: Springer.
50. Lehmann, H., Nagai, Y., & Metta, G. (2016). The question of cultural sensitive gesture libraries in HRI—An Italian—Japanese comparison. In *Proceedings of ICDL-Epi 2016.*
51. Tanner, J. E., & Byrne, R. W. (1996). Representation of action through iconic gesture in a captive lowland gorilla. *Current Anthropology, 37*(1), 162–173.
52. Trovato, G., Zecca, M., Sessa, S., Jamone, L., Ham, J., Hashimoto, K., et al. (2013). Cross-cultural study on human-robot greeting interaction: Acceptance and discomfort by Egyptians and Japanese. *Paladyn, Journal of Behavioral Robotics, 4*(2), 83–93.
53. Trovato, G., Zecca, M., Do, M., Terlemez, Ö., Kuramochi, M., Waibel, A., et al. (2015). A novel greeting selection system for a culture-adaptive humanoid robot. *International Journal of Advanced Robotic Systems, 12*(4), 34.

54. Rossi, P. G. (2011). Didattica enattiva. *Complessità, teorie dell'azione, professionalità docente: Complessità, teorie dell'azione, professionalità docente*. FrancoAngeli.

55. Lau, K. W., Tan, H. K., Erwin, B. T., & Petrovic, P. (1999). Creative learning in school with LEGO (R) programmable robotics products. In *29th Annual IEEE Frontiers in Education Conference* (Vol. 2, pp. 12–26).

56. Balogh, R. (2010). Educational robotic platform based on arduino. In *Proceedings of Conference on Educational Robotics* (pp. 119–122).

57. Mukai, H., & McGregor, N. (2004). Robot control instruction for eighth graders. *IEEE, Control Systems, 24*(5), 20–23.

58. Papert, S., & Harel, I. (1991). Situating constructionism. *Constructionism, 36*(2), 1–11.

59. Hirst, A. J., Johnson, J., Petre, M., Price, B. A., & Richards, M. (2003). What is the best programming environment/language for teaching robotics using Lego Mindstorms? *Artificial Life and Robotics, 7*(3), 124–131.

60. Powers, K., Gross, P., Cooper, S., McNally, M., Goldman, K. J., Proulx, V., et al. (2006). Tools for teaching introductory programming: What works? *ACM SIGCSE Bulletin, 38*(1), 560–561.

61. Bell, S. (2010). Project-based learning for the 21st century: Skills for the future. *The Clearing House, 83*(2), 39–43.

62. Haas, M. D., Baxter, P., de Jong, C., Krahmer, E., & Vogt, P. (2017, March). Exploring different types of feedback in preschooler and robot interaction. In *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction* (pp. 127–128).

63. Belpaeme, T., Kennedy, J., Baxter, P., Vogt, P., Krahmer, E. E., Kopp, S., et al. (2015). L2TOR-second language tutoring using social robots. In *Proceedings of the ICSR 2015 WONDER Workshop*.

64. Ahmad, M. I., Mubin, O., Shahid, S., & Orlando, J. (2019). Robot's adaptive emotional feedback sustains children's social engagement and promotes their vocabulary learning: A long-term child–robot interaction study. *Adaptive Behavior, 27*(4), 243–266.

65. Mubin, O., Stevens, C. J., Shahid, S., Al Mahmud, A., & Dong, J. (2013). A review of the applicability of robots in education. *Technology for Education and Learning, 1,* 1–7.

66. Tanaka, F., Isshiki, K., Takahashi, F., Uekusa, M., Sei, R., & Hayashi, K. (2015). Pepper learns together with children: Development of an educational application. In *IEEE-RAS 15th International Conference on Humanoid Robots* (pp. 270–275).

67. Ishiguro, H., Ono, T., Imai, M., Maeda, T., Kanda, T., & Nakatsu, R. (2001). Robovie: An interactive humanoid robot. *Industrial Robot: An International Journal, 28*(6), 498–504.

68. Han, J., & Kim, D. (2009) r-Learning services for elementary school students with a teaching assistant robot. In *4th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 255–256). IEEE.

69. Shamsuddin, S., Ismail, L. I., Yussof, H., Zahari, N. I., Bahari, S., Hashim, H., & Jaffar, A. (2011) Humanoid robot NAO: Review of control and motion exploration. In *2011 IEEE International Conference on Control System, Computing and Engineering (ICCSCE)* (pp. 511–516).

70. Belpaeme, T., Kennedy, J., Ramachandran, A., Scassellati, B., & Tanaka, F. (2018). Social robots for education: A review. *Science Robotics*, *3*(21), eaat5954.

71. Tanaka, F., & Kimura, T. (2009) The use of robots in early education: A scenario based on ethical consideration. In *The 18th IEEE International Symposium on Robot and Human Interactive Communication. RO-MAN 2009* (pp. 558–560). IEEE.

72. Tanaka, F., & Matsuzoe, S. (2012). Children teach a care-receiving robot to promote their learning: Field experiments in a classroom for vocabulary learning. *Journal of Human-Robot Interaction, 1*(1), 78–95.

73. Leite, I., Castellano, G., Pereira, A., Martinho, C., & Paiva, A. (2012). Modelling empathic behaviour in a robotic game companion for children: An ethnographic study in real-world settings. In *Proceedings of 7th ACM/IEEE International Conference on HRI* (pp. 367–374). ACM.

74. Dautenhahn, K., Nehaniv, C. L., Walters, M. L., Robins, B., Kose-Bagci, H., Mirza, N. A., et al. (2009). KASPAR—A minimally expressive humanoid robot for human–robot interaction research. *Applied Bionics and Biomechanics, 6*(3–4), 369–397.

75. Iacono, I., Lehmann, H., Marti, P., Robins, B., &. Dautenhahn, K. (2011) Robots as social mediators for children with Autism—A preliminary analysis comparing two different robotic platforms. In *2011 IEEE International Conference on Development and Learning (ICDL)* (Vol. 2, pp. 1–6). IEEE.

76. Köse, H., Akalin, N., & Uluer, P. (2014). Socially interactive robotic platforms as sign language tutors. *International Journal of Humanoid Robotics, 11*(01), 1450003.

77. Köse, H., Uluer, P., Akalın, N., Yorgancı, R., Özkul, A., & Ince, G. (2015). The effect of embodiment in sign language tutoring with assistive humanoid robots. *International Journal of Social Robotics, 7*(4), 537–548.

78. Uluer, P., Akalın, N., & Köse, H. (2015). A new robotic platform for sign language tutoring. *International Journal of Social Robotics, 7*(5), 571–585.

79. Kanda, T., Hirano, T., Eaton, D., & Ishiguro, H. (2004). Interactive robots as social partners and peer tutors for children: A field trial. *Human-Computer Interaction, 19*(1–2), 61–84.

80. Gorostiza, J. F., Barber, R., Khamis, A. M., Malfaz, M., Pacheco, R., Rivas, R., & Salichs, M. A. (2006). Multimodal human-robot interaction framework for a personal robot. In *The 15th IEEE International Symposium on Robot and Human Interactive Communication. ROMAN 2006* (pp. 39–44). IEEE.

81. Benitti, F. B. V. (2012). Exploring the educational potential of robotics in schools: A systematic review. *Computers and Education, 58*, 978–988.

82. Benitti, F. B. V., & Spolaôr, N. (2017). How have robots supported stem teaching? In *Robotics in STEM Education* (pp. 103–129). Cham: Springer.

83. Lehmann, H., & Rossi, P. G. (2019a). Enactive Robot Assisted Didactics (ERAD): The role of the maker movement. In *Proceedings of the International Conference on Educational Robotics*.

84. Damiano, E. (2013) La mediazione didattica. *Per una teoria dell'insegnamento: Per una teoria dell'insegnamento*. FrancoAngeli.

85. Laurillard, D. (2012). Teaching as a design science. *Building pedagogical patterns for learning and technology*.

86. Lehmann, H. & Rossi, P. G. (2019b). Social robots in educational contexts: Developing an application in enactive didactics. *Journal of e-Learning and Knowledge Society 15*(2).

87. Terlemez, Ö., Ulbrich, S., Mandery, C., Do, M., Vahrenkamp, N., & Asfour, T. (2014). Master Motor Map (MMM)—Framework and toolkit for capturing, representing, and reproducing human motion on humanoid robots. In *2014 IEEE-RAS International Conference on Humanoid Robots* (pp. 894–901). IEEE.