Socially Inclusive Robots: Learning Culture-Related Gestures by Playing with Pepper

Berardina De Carolis[a], Francesca D'Errico[c], Nicola Macchiarulo[a] and Giuseppe Palestra[c]

[a]Department of Computer Science, University of Bari, Bari, Italy
[b]ForPsiCom Department, University of Bari, Bari, Italy
[c]Hero S.r.l., Martina Franca, Italy

Abstract

Social robots are being used successfully as educational technologies. In this paper we propose to use a social robot, Pepper in this case, to teach and explain the meaning of culture-related gestures to unaccompanied minor migrants and support their integration in a new culture. The use of social humanoid robots seems to be an adequate interaction mean to this aim, since they can provide both examples of gesture executions, explanations about the meaning and the context in which the gesture should be used. Moreover, as in other assistive domains, social robots may be used to attract the children's attention and support the social operator in establishing contact with these children that very often, after the difficulties of the journey, do not trust adults. Results of a preliminary study, even if performed with a limited number of participants, show that the proposed approach is suitable for learning gestures.

Keywords

Social robot, gestures recognition, cultural gestures, serious game

1. Introduction

In the past years, numerous migrant children, often unaccompanied, applied for international protection in EU countries. EU has the responsibility of supporting these children and, at the same time, the chance to nurture their potential to enhance their contribution to our societies. Even if, in this period of the pandemic, the number of migrants decreased, it is necessary to design solutions not only to ensure their access to care and education but also to foster their social inclusion [1]. In this context, is of particular importance to teach the language of the country in which they are hosted and also how to interpret and use culture-related gestures since human communication is multimodal [2] and, according to some studies and in some cultures, the vocabulary of hand gestures is very rich especially in Italy [3]. Gestures are used to convey the meaning of a message [4]. There are several types of gestures: metaphoric gestures (i.e., those that explain a concept), deictic gestures (i.e., pointing movements), and iconic gestures among others. However, many gestures are culture-dependent and do not have a unique meaning and symbolism. The same gesture can mean something quite nasty and disrespectful to a person from a different cultural background. Hand gestures are a very important part when learning a foreign language [5]. Also, when migrant children arrive in a country, after the migration process, they generally face a so-called 'acculturative stress [6] and thus they can feel to be not accepted by the host country. Social Robots are embodied autonomous intelligent entities that interact with people in everyday environments, following social behaviors typical of humans [7, 8]. Social robots are mainly used to improve people's experience in several application domains, language teaching among others. The work of [9, 10] show that children who are taught by a robot as opposed to a human teacher store new words of a second language faster and better in their long-term memory. Moreover, social robots are less complex and less intimidating than humans and may provide effective support during triadic therapy or intervention. They may be programmed to have a deterministic behavior that can be repeated as many times as needed. In the first phase of this project, we developed first an application, NaoKi, that used the Nao robot for teaching Italian and culture-dependent gestures to migrants [11]. Based on the results of a study made with this prototype of the system, we updated the developed solution for running on the Pepper robot. Pepper, due to its characteristics, is suitable for the application context for several reasons, one of these is represented by the possibility to synchronize what the robot says with the display of useful information of the tablet (feedback, images, examples) for making the comprehension of the explanation easier. As in the previous approach, to recognize gestures in real-time, we used Kinect, a device able to detect and the user's skeleton, that has been integrated with the Pepper robot. An important component of the application is represented by the gesture database that has been

Joint Proceedings of the ACM IUI 2021 Workshops, April 13–17, 2021, College Station, USA
berardina.decarolis@uniba.it (B. D. Carolis); francesca.derrico@uniba.it (F. D'Errico); nicola.macchiarulo@uniba.it (N. Macchiarulo); giuseppepalestra@gmail.it (G. Palestra)

designed using the formalization proposed by [3]. Even if, due to the COVID-19 emergency, it was not possible to perform a large study with migrant children, results obtained with the use of Pepper are in line with those obtained with Nao. However, from the observation of the interaction and the answers to some of the questions in the post-experiment interview, the user experience and the believability of the interaction with the robot seem to be better evaluated in the case of Pepper than of the Nao robot. The paper is structured as follows. Section 2 of the paper briefly describes the background and motivations of using a social robot in assistive domains, with emphasis on issues concerning the reception of migrant children. Then, we briefly explain the structure of the Italian Gestionary and then, in Section 4, how it has been used in the PepperKi application. Then we report the results of a preliminary evaluation study in Section 5. Conclusions and future work directions are discussed in the last section.

2. Background and Motivation

In the last years Italy, along with other European countries, has become the landing place for numerous migrant children often unaccompanied (Unaccompanied Foreign Minors - UFM -, isolated children, separated children). After the first reception phase in communities for minors without a family, the second level reception of the Protection System SIPROIMI – Protection System for International Protection Holders and Unaccompanied Foreign Minors - has the aim of ensuring to children the living conditions appropriate to their age, with interventions aimed at social integration and autonomy. Then, the minor will be assigned, through the Juvenile Court, to a tutor. Due to their age, these children live a double status: that of a minor and a migrant, experiencing both the difficulties related to the abrupt and rapid passage to adulthood and the integration into a new and different society [6]. They have profound traumas due to many difficulties encountered along the way to reach Italy (repeated beatings, threats, abuses, hunger and thirst). Moreover, they suffer of the cultural shock and abandonment of parental figures of reference, essential in the life of any child. For these reasons, the minors may show distrust towards the operators and any physical contact, even a simple hand on the shoulder, can generate a sense of anxiety in the child. Reception activities, specifically designed by the government, are implemented to offer psychological support following the disastrous journey. Among the various activities, whose objectives have been defined based on the characteristic of the target group, there are some (the BLUE and GREY activities) aiming to share key words about greetings and moods (in Italian) and to know customs and cultural habits of the country of origin, respectively. At this stage, a social robot could play an important supporting role of cultural mediator, who have the important task of trying to understand the child's expectations and lived experience, to transmit information about integration into Italian society, then acting as a bridge between the two cultures. However, one of the major difficulties of the operators consists in attracting and keeping the attention of the children that most of the times, is very difficult since they are scared and may not accept physical contact. Most of the young migrants have their gaze down, indicating their emotional situation and their lack of self-awareness, they speak with a low voice and have difficulty in looking at operators. The robot could be of great help both for the operators so that they can establish a relationship with the minors through the interactions with the robot as a "game". In this way, minors could learn the first necessary information to start the integration process that, initially is based primarily on the knowledge of the Italian language, which must already be learned in the first reception center. A social robot is a physically embodied, autonomous agent that communicates and interacts with humans at a social and emotional level. The robot should be able to interpret properly human behavior, to react to changes during the interaction in a socially plausible manner. The use of social robots in education has been shown to be successful in diverse contexts [8, 12]. In particular, the use of robots may increase attention, engagement, and compliance, which are critical components of successful learning [13]. Social robots are employed in education with particular success with children with special needs. For instance, social robots are widely applied to teach basic social skills to children with autism, since they resemble humans but are less complex, seem to be able to manage these issues successfully [14, 15, 16]. As far as the efficacy of using social robots in the teaching of a foreign language and culture is concerned, recent research shows that it may lead to interesting results [17]. A social robot was employed successfully as an assistant to teach English vocabulary to Iranian students [10]. Also, the robot-assisted group showed improved retention of the acquired vocabulary. Other studies suggest that the sociality of the robot improved the learning outcomes [18, 19, 20]. Then, robots open up new possibilities in teaching that were previously unavailable, leaving space to explore novel aspects of language learning, as culture-dependent gestures. Gesture recognition in Human-Robot Interaction has been proposed in [21] to allow people, especially those with physical limitations, to give instructions to the robot easily and intuitively. The system uses a Kinect for gesture detection, and recognition is performed using a Microsoft software, Visual Gesture Builder. Research, similar to the one described in this paper, has been conducted on children with autism spectrum disorders [22]. Since these children have delayed gestural development,

Table 1
An example of gesture description (Poggi 2006).

Gesture	Verbal Formulation	Hand Shape	Orientation	Location	Movement	Non-manual Components
	**I pray you**	Hands open with palms touching	Fingers pointing towards the chin	Neutral space	United hands moving slowly forward and backward always keeping your fingers pointing upwards	

a social robot was used to teach them to recognize and produce eight pantomime gestures that expressed feelings and needs. This study reports that children in the intervention group were able to recognize more gestures and generalize the acquired gestural recognition skills to human-to-human interaction. Also, in [23] the social robot Nao has been used in conjunction with Kinect for developing a serious game for sign language tutoring. All these studies report how a social robot represents a successful interface for teaching a second language and, in teaching gestures to children, especially to those with special needs. Looking at the activities planned for migrant children, the robot, through games, images, and sounds, can facilitate the acquisition of the first words in Italian, useful to be understood in Italy. This could be integrated with lessons from the robot on the Italian culture and on that of the different countries of origin, to analyze the two cultures, highlighting differences and elements in common. Finally, in this phase, the robot can be fundamental to facilitate the process of understanding the gestures typical of Italian culture that favor their social integration.

3. The Gesture Dictionary

Communication is an intentional process of sending, receiving and interpreting messages conveyed through verbal or nonverbal signals [24]. In verbal communication, the goal of communicating can be almost intentional, instead, when implemented with other types of non-verbal signals such as, for example, gestures, the goal can more likely, than in verbal case, defined as unconscious or tacit. The correspondences between signals of our body (gestures, expressions) and meanings are different from culture to culture even if, in many communication systems, such as facial expressions and gaze, they are almost universal [25]. We define ‘gesture’ any movement made with hands, arms and/or shoulders. In particular, the hands play a significant role in communication and can be classified in [4, 26, 27]:

1. *Deictic*, which indicates an object or a person (i.e. pointing with the index finger);
2. *Iconic*, which depicts, in the air, the form or imitate the movements of an object, an animal, a person;
3. *Batonic*, in which the hands move rhythmically from the top to the bottom to scan and highlight the accented syllables in a sentence;
4. *Symbolic*, gestures that, in a given culture, have a shared meaning, and thus culture-dependent since they are learned from an early age by observing them daily, then they are often incomprehensible to people of different cultures. This latter typology has particular relevance in the application context of this work.

When we talk about communicative gestures, which correspond to a signal-meaning pair, it is necessary, for those who have different cultures, to have a ”dictionary” of such gestures, to allow a better translation of the communicative act. To this aim, [3] proposed the ”Italian Gestionary”, a useful resource in which each gesture is presented through a picture, a description of the movement, and the corresponding verbal description (Table 1). Gestures can be considered as a piece of semantic information present in the mind of those who want to communicate. It is possible to extract the meaning and use of a gesture by analyzing the Gestionary that, for each gesture, provides:

a. a *verbal formulation*, the associated verbal meaning, expressed using a real sentence, for example, the gesture of applauding can be paraphrased as ”compliments!”;
b. a *context*, that can contribute to the core meaning [24];
c. a *meaning* is the core definition across different contexts, similar to that of word dictionary;
d. a *grammatical classification*, in terms of ”gestures-sentences”, called holophrastic, from ”word gestures”, called articulated, depending on that have the meaning of a whole sentence or only a part;
e. a *pragmatic classification* that concerns only the ”phrase-gestures” which are also classified accord-

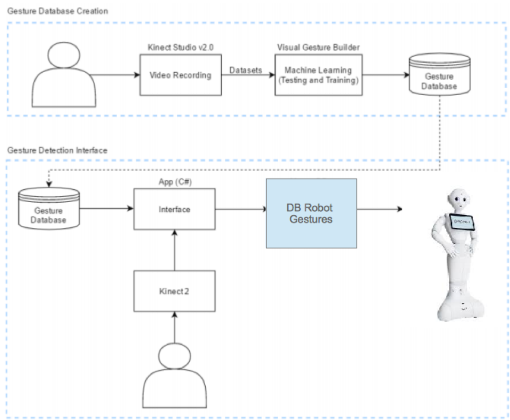

Figure 1: The architecture scheme of the PepperKi application.

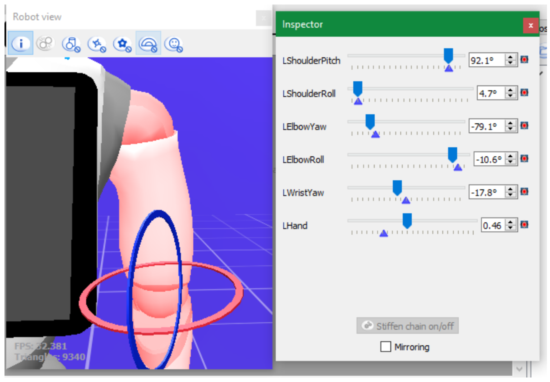

Figure 2: An example of the "Timeline Script" used to build the robot's gestures.

ing to their specific performative just like the gesture of praise (the applause, "Congratulations!");

f. a *semantic classification* that provides information on the world, like among others the times ("yesterday", "after") or quantities ("two") or also information about the mind, emotions, and degree of knowledge of the person with whom you are communicating;

g. a *rhetorical meaning*, that is a rhetorical use of the gesture different from the literal one.

For example, Poggi [3] describes the gesture of hitting the chest with the hand, with the palm facing downwards and the fingers touching by providing as meaning "I do not digest it", but in a rhetorical way because what is not digested is not a food but a person who metaphorically "can't stand". In a gesture identification and representation is important to classify not only the verbal formulation but also their description in terms of shape, orientation in the space and type of movement since they synchronously contribute to the final meaning as represented in Table 1 [3].

4. The Application

In order to execute, recognize and explain the meaning of the gestures using a social robot, we integrated the Pepper robot with Microsoft Kinect through the PepperKi application (Figure 1). The proposed architecture has been described in [11] in which we used Nao as social robot. In this case we used the same architecture but with Pepper. As a first step, Kinect is used to build the database of gestures to be successively recognized during the learning session with Pepper. Kinect for Windows is a device that allows to recognize and track the body (via the skeletal tracking function) of one or more people (up to 6). The Kinect SDK 2.0 is a library that is essential for recording the video of the gesture to be recognized and, then, converting and analyzing it with the Visual Gesture Builder. Using this technology, the application allows the creation of a database of gestures that can be recognized and the degree of correctness of the recognized gesture with respect to the selected one.

The second component, the Gesture Detection Interface, allows connecting the application to the Pepper robot that is used, in this case, to interact with the children in the learning task of gestures. To do so we enriched the robot gesture database with the coordinates of the gesture to allow the robot to play it. To do so, we used the functionality of the robot programming environment (Choregraphe) that allows to memorize a new movement using the "Timeline" scripts (Figure 2). Whit this functionality just moving the robot's body parts and then clicking on "save" it is possible to save the gesture and the behavior and then re-run it when selected.

The database is running in the background of both components and contains the description of the gestures. It has been structured according to the Gestionary presented in Section 3 with, in addition, the coordinates of the gestures recorded by the Kinect SDK and the description of the gesture as programmed in Pepper. In particular, we stored the movement coordinates of static/discrete (i.e. hand on the forehead) and continuous/dynamic gestures (i.e. waiving to say hello). Another important functionality regards the check of the correctness of the performed gesture. To allow this, we need to query the ".gbd" database (DB Robot Gesture in Figure 1) containing the coordinates of the gestures recognized by the Kinect; the coordinates are searched based on the gesture and then the curve of correctness and finally, the evaluation is done. In the database, together with the information about gesture execution, we linked its description and a scenario to be used by the robot to better explain the context in which the gesture is used. Gestures implemented so far are used to express:

- **cognitive states** like *"I am thinking"* (hand and index finger touching the chin or the forehead), *"I don't believe it"* (hand with the dorso toward the forehead and the head looking down), *"I don't understand what you are saying"* (pinched fingers);
- **emotional states** like *"I pray you"* (hold hands together moving them slowly forward and backward and with the head down, see Table 1),*"I love you/kissing"*, (hold one or both hands close the mouth, and move them in the direction of the interlocutor with the palm raised);
- **regulation** of the interaction with an interlocutor like *"Greeting/Hello"* (waving the hand), *"Ask to speak"* (rising the right hand).

The application offers two *learning modalities* in which the interaction may be initiated by the robot or the child. In the first modality, the robot shows a gesture to the child, explains the meaning of the gesture, by taking it from the database, and asks the child to reproduce it. In the example in Figure 3, Pepper teaches a gesture to a child. In this case, the robot is teaching the gesture *"I pray you"*. First, the robot shows how to execute the gesture and then asks the child to reproduce it. In the second modality, the child asks the robot to perform a specific gesture or a gesture conveying a specific meaning. Moreover, the child may ask for an explanation about the gesture's meaning. When the child performs the gesture, the robot gives him feedback that is generated according to the correctness score calculated by Kinect. In case of correct execution, Pepper enjoys and shows a green star on the tablet otherwise it encourages the child to try again, shows again the gesture and asks to repeat it.

(a) "I pray you" – Looking at the gesture

(b) "I pray you" – Repeating the gesture

Figure 3: (a) Pepper shows the child how to perform "I pray you", (b) the child repeats the gesture.

5. Evaluation

We performed the same exploratory experiment with the new version of the game with 10 Italian children with an age from 6 to 10 y.o. Unluckily, due to the pandemic, it was not possible to involve migrant children. We asked each subject to learn and perform the following gestures three times: "I pray you", "Hello", "Kissing", "I don't understand". Before the experiment, children gave their verbal informed consent and their parents provided written informed consent for participation in the study. At the end of the interaction, we asked subjects to answer a simple survey about their experience. The survey was composed of six statements and participants were asked to evaluate each of them on a Likert scale from 1 to 5. Some statements concerned the evaluation of the interaction - if easy and engaging - in general, some others were specifically concerning gesture execution and recognition. The statements were the following:

- Q1: I was able to interact with the robot
- Q2: It was easy to understand how to perform gestures
- Q3: It was easy to understand the feedback
- Q4: The system had an adequate response time
- Q5: It had a low number of misidentified gestures
- Q6: Interacting with the robot was engaging

As far as task completion is concerned, each subject was able to complete the tasks and to perform the required gesture. Figure 4 shows the results of the survey compared to the previous study with Nao.

The descriptive means are higher than 4 and thus they indicate that the interaction was quite satisfactory both from the engagement and gesture learning points of view, mainly because they are easy to understand and the participants were Italian. Even if the study was conducted in both cases with a limited number of participants and was not conducted with migrants, the results with Pepper are higher on average compared to those obtained in the previous evaluation with Nao. To confirm this, we performed the Mann-Whitney Test and the resulting p-value was 0.00058 which is significant at $p < .05$, thus showing that the interaction with Pepper was better evaluated than the one with Nao for the learning gesture task.

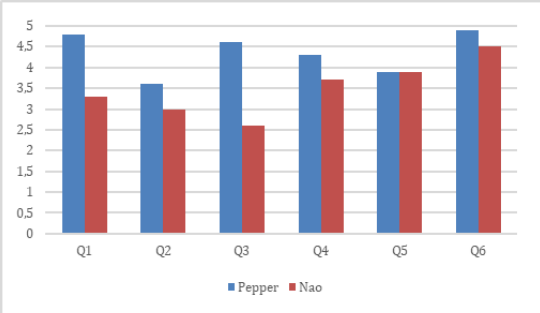

Figure 4: Comparison of the means for each question of the survey in the two studies.

6. Conclusions and Future Work Directions

In this paper, we approached the problem of teaching culture-related gestures to migrant children for improving their social integration in a different country and culture using the social robot Pepper. To this aim, we followed the same approach proposed in [11] in which Microsoft Kinect was integrated with the Nao robot. In this new study, we modified the system by using Pepper instead of Nao and adding visual feedback on its tablet. Results of the evaluation of the new prototype seem to confirm the previous ones. Pepper was better evaluated than Nao probably because it has almost the same height as the children that could better mirror the proposed gesture. However, we cannot say that the reason is related only to the robot body since we did not control the language knowledge and their potential difficulties. In both cases, as in other assistive domains, the robot has the capability of attracting attention, and then in the future, it will be used to support the social operator in establishing contact with these children. We are aware that the main limitation of this work concerns the number and the categories of subjects involved in the evaluation study. Even if there were not migrant children, the preliminary results show the feasibility and efficacy of the approach. Future work will take into account children's basic linguistic knowledge and other socio-psychological variables, also associated with the children's sense of mastery during the interaction. Moreover, we will implement a larger sample of gestures compatible with Pepper's capabilities. In particular, we will see if the approach is suitable for teaching to Italian children new gestures that are not typical of the Italian culture for improving cultural integration from both sides.

References

[1] C. Penna, Apprendimento sinergico innovativo. Percorsi educativi per minori stranieri, IN-CON-TRA.: Didattica e Pedagogia dell'inclusione, Aracne, 2014.
[2] I. Poggi, Mind, Hands, Face and Body: A Goal and Belief View of Multimodal Communication, Weidler, 2007.
[3] I. Poggi, Le parole del corpo: introduzione alla comunicazione multimodale, Carocci, 2006.
[4] D. McNeill, Hand and Mind: What Gestures Reveal about Thought, University of Chicago Press, 1992.
[5] P. Diadori, Gestualita e insegnamento linguistico (gestures and language teaching), Rassegna Italiana di Linguistica Applicata (1987) 33–50.
[6] K. Lueck, M. Wilson, Acculturative stress in latino immigrants: The impact of social, socio-psychological and migration-related factors, International Journal of Intercultural Relations 35 (2011) 186–195.
[7] A. Billard, K. Dautenhahn, Grounding communication in autonomous robots: An experimental study, Robotics and Autonomous Systems 24 (1998) 71–79. Scientific Methods in Mobile Robotics.
[8] T. Fong, I. Nourbakhsh, K. Dautenhahn, A survey of socially interactive robots, Robotics and Autonomous Systems 42 (2003) 143–166. doi:10.1016/S0921-8890(02)00372-X.
[9] T. Schodde, K. Bergmann, S. Kopp, Adaptive robot language tutoring based on bayesian knowledge tracing and predictive decision-making, in: Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction, HRI '17, Association for Computing Machinery, New York, NY, USA, 2017, p. 128–136.
[10] M. Alemi, A. Meghdari, N. M. Basiri, A. Taheri, The effect of applying humanoid robots as teacher assistants to help iranian autistic pupils learn english as a foreign language, in: A. Tapus, E. André, J.-C. Martin, F. Ferland, M. Ammi (Eds.), Social Robotics, Springer International Publishing, Cham, 2015, pp. 1–10.
[11] B. De Carolis, G. Palestra, C. Penna, M. Cianciotta, A. Cervelione, Social robots supporting the inclusion of unaccompanied migrant children: Teaching the meaning of culture-related gestures, Journal of E-Learning and Knowledge Society 15 (2019).
[12] J. Kennedy, P. Baxter, E. Senft, T. Belpaeme, Higher nonverbal immediacy leads to greater learning gains in child-robot tutoring interactions, 2015.
[13] A. Ramachandran, C.-M. Huang, E. Gartland, B. Scassellati, Thinking aloud with a tutoring robot to enhance learning, in: Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction, HRI '18, Association for Com-

puting Machinery, New York, NY, USA, 2018, p. 59–68.

[14] G. Palestra, B. D. Carolis, F. Esposito, Artificial intelligence for robot-assisted treatment of autism, in: WAIAH@AI*IA, 2017.

[15] V. Pennazio, Social robotics to help children with autism in their interactions through imitation, Research on Education and Media 9 (2017).

[16] A. Duquette, F. Michaud, H. Mercier, Exploring the use of a mobile robot as an imitation agent with children with low-functioning autism, Auton. Robots 24 (2008) 147–157.

[17] L. Boccanfuso, S. Scarborough, R. Abramson, A. Hall, H. Wright, J. O'Kane, A low-cost socially assistive robot and robot-assisted intervention for children with autism spectrum disorder: field trials and lessons learned, Autonomous Robots 41 (2017).

[18] T. Kanda, T. Hirano, D. Eaton, H. Ishiguro, Interactive robots as social partners and peer tutors for children: A field trial, Human Computer Interaction (Special issues on human-robot interaction) 19 (2004) 61–84.

[19] M. Saerbeck, T. Schut, C. Bartneck, M. Janse, Expressive robots in education varying the degree of social supportive behavior of a robotic tutor, volume 3, 2010, pp. 1613–1622.

[20] F. Tanaka, S. Matsuzoe, Children teach a care-receiving robot to promote their learning: Field experiments in a classroom for vocabulary learning, Journal of Human-Robot Interaction 1 (2012).

[21] H. P.F.P., Gesture recognition with microsoft kinect tools for socially assistive robotics scenarios. master thesis., ???? URL: https://fenix.tecnico.ulisboa.pt/downloadFile/1970719973966417/tese-robot-coach%20(1).pdf.

[22] W. C. So, M. Wong, W. Lam, C. Lam, D. Fok, Using a social robot to teach gestural recognition and production in children with autism spectrum disorders, Disability and Rehabilitation Assistive Technology 13 (2018).

[23] H. Kose, R. Yorganci, I. I. Itauma, Humanoid robot assisted interactive sign language tutoring game, in: 2011 IEEE International Conference on Robotics and Biomimetics, 2011, pp. 2247–2248.

[24] I. Poggi, F. D'Errico, Social signals: A framework in terms of goals and beliefs, Cognitive processing 13 Suppl 2 (2012) 427–45.

[25] P. Ekman, Facial action coding system (facs), A Human Face (2002).

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

[27] A. Kendon, Gesture: Visible action as utterance (2004).