

How to Manage Affective State in Child-Robot Tutoring Interactions?

Thorsten Schodde
CITEC, Bielefeld University
Bielefeld, Germany
tschodde@techfak.uni-bielefeld.de

Laura Hoffmann
CITEC, Bielefeld University
Bielefeld, Germany
lahoffmann@techfak.uni-bielefeld.de

Stefan Kopp
CITEC, Bielefeld University
Bielefeld, Germany
skopp@techfak.uni-bielefeld.de

Abstract—Social robots represent a fruitful enhancement of intelligent tutoring systems that can be used for one-to-one tutoring. The role of affective states during learning has so far only scarcely been considered in such systems, because it is unclear which cues should be tracked, how they should be interpreted, and how the system should react to them. Therefore, we conducted expert interviews with preschool teachers, and based on these results suggest a conceptual model for tracing and managing the affective state of preschool children during robot-child tutoring.

I. INTRODUCTION

The use of robots for educational purposes has increasingly moved into focus in recent years. One rationale is to enable individually adapted one-to-one teaching for weaker students, which can hardly be provided in regular classrooms. This idea already underlay educational on-screen applications like intelligent tutoring systems (ITSs). Physically present social robots are expected to bring an additional quality to the learning interactions, similar to co-present teacher-child or child-child interaction, which can make the tutoring experience more effective. Indeed, a recent study showed that students' learning performance increases up to 50% if a social robot was included compared to a classical on-screen media learning [1].

One of the main challenges for robot tutors is to identify the learner's internal states, e.g., whether she is following, distracted, or losing motivation. Yet, recognizing and reacting to these cognitive and affective states is vital to keep the learner engaged and to foster learning. In previous work, we developed an approach to dynamically adapt robot tutoring to the changing pedagogical state of the learner [2]. There, the skill mastery of the student is kept track of inferentially using Bayesian Knowledge Tracing, which enables the robotic tutor to choose the to-be-addressed skill and difficulty of the next task accordingly. This way the model works to keep the child in the "zone of proximal development" [3], which can lead to a feeling of flow, motivation and better learning [4], [5].

However, this approach lacks "emotional intelligence" [6]. Successful human teachers not only teach the curriculum according to the learner's knowledge state, but also manage the affective states of children. Studies have shown that affective states like curiosity, interest, flow, joy, boredom, frustration and surprise can influence learner's problem-solving abilities, and affect task engagement and learning motivation [7]. Further, such states are found to influence cognitive processes like

long-term memorizing, attention, understanding, remembering, reasoning, decision-making and the application of knowledge in task solving [4], [8]. It is thus not surprising that good human tutors are sensitive to learners' vocal (e.g., intonation) or nonvocal behavior (e.g., facial expression, body language) [9]. Technical systems are also increasingly able to recognize most of these cues - albeit sometimes in a quite rudimentary way. However, little attention has been paid to the question how a robot should interpret and respond to the affective state of a learner during tutoring with the needed flexibility and adaptiveness [10], [11].

In this paper we present steps towards a model for tracing and managing the affective state of preschool children in second language tutoring interactions with a robot tutor. This model is based on pedagogical knowledge about children's affective states during actual robot-child tutoring gathered through expert interviews with preschool teachers. This knowledge comprises information about which affective states are relevant, from which features they can be tracked and, finally, how to react to them appropriately as a tutor. It lends itself to a decision-theoretic affective state tracing model that can be combined with our previously developed adaptive knowledge tracing approach. The following section discusses previous work on affect detection and affective tutoring systems. Afterwards we present the procedure and results of the conducted expert interviews. Finally, we discuss how these findings can be incorporated into a conceptual model that enables the recognition of and reaction to changes in children's affective states.

II. RELATED WORK

A. Affect Detection

A lot of work has been done on affect recognition based on different modalities. One widely used approach is the analysis of facial expressions to detect the affective state of a user [12]. Often, classifiers are trained on "very expressive and played" emotions, making their applicability to real-world interactions questionable. In fact, the accuracy of emotion detection based on facial features is often low in real-world applications. Furthermore, the recognition rate is strongly dependent on the expressiveness of each target.

An alternative approach is the detection of affect from the user's voice [13]. Classifiers based on voice analysis are

trained on datasets of spontaneous speech, so that they are more suitable for real-world applications. With regard to robot-child tutoring, affect detection through speech analysis is, however, difficult because speech input is often not included as speech recognition for children has a low accuracy [14]. Other attempts have been made to detect the affective state through analyzing written text [15]. This approach includes, for instance, analyzing the usage of adjectives and adverbs. But in most natural interactions humans do not write text, and preschool children are usually not able to read and write.

A broader approach for affective state detection is the tracking of the whole body posture and movements by using a body pressure mat laying on a seat [16], or using a Microsoft Kinect [17]. A limitation is that the use of a body pressure mat assumes that the user remains on a seat and cannot move around. The Kinect, however, allows the user to move around, but may have problems in detecting smaller events like small postural shifts. Also, approaches based on human physiology have been adopted. In this realm, measures such as ECG, EEG, EMG [18], [19], and brain imaging [20] have been applied to “read” the affective state from the user’s body. The results of these methods are promising, however the applicability of such obtrusive approaches (e.g., wires and patches on the body) in tutoring interactions with children is clearly limited.

In sum, all of these approaches have their field of use, but also their limitations. In contrast, multi-modal approaches have been studied to overcome these limitations and to increase accuracy of the detection. A lot of combinations exist, e.g., facial expressions and voice [21], facial expression, voice and body posture [22], facial expressions, body postures and context dependent activity logs [23], or speech and text [24]. Such systems demonstrated that a multi-modal approach to detect affective states results in higher accuracy rates.

B. Affective Tutoring Systems (ATs)

Since the technical progress yields new possibilities to make use of the affective state in tutoring interactions, a lot of systems have been extended with such a module. Shen et al. [25], for instance, used physiological signals for affect detection and then guided the learning interaction by different affective strategies. Their results demonstrated the superiority of an emotion-aware over a non-emotion-aware system with a performance increase of 91% .

Alexander et al. [26] developed an affect-detecting ITS including a virtual agent for primary school students. The affective state is detected by analyzing the facial expressions of the student and serves as the basis for a case-based selection of the next tutoring actions. The case-based rules have been informed by an observational study of human tutors. In a study conducted in a primary school, where children had to solve mathematical equations, the use of their affective system showed a significant increase of the students’ performance as compared to a control group without affective support.

The “Affective AutoTutor” system [27] can automatically detect boredom, confusion, frustration and neutral affect by monitoring conversational cues and discourse features along

with gross body language and facial features. Cues provided by each channel are combined to select a single affective state, based on which AutoTutor responds with empathic, motivational, or encouraging dialog-moves and emotional displays. Evaluations showed that this systems is able to support learners not only in acquiring knowledge, but also in using it in transfer tasks later on. Recently, Goren et al. [28] incorporated affect detection via facial expressions in robot-child tutoring. In a study with preschool children they showed that their system personalized its policy over the course of training, and that children who interacted with the personalized robot showed increased long-term positive valence as compared to a control group without personalization.

Taken together, the findings from earlier approaches suggest the inclusion of affect detection in robot-child tutoring. Most affect detectors are trained on specifically annotated data to identify the important cues for each affective state. For example, the emotion classifier “Affectiva Affdex” [29] is trained on more than 5 million human faces to classify facial expressions. Strategies for how to respond to those states are usually based on observational studies of the reactions of a human tutor to the behavior of a student [30]. We adopt this approach here, too, with the aim of building a model that enables a robot to detect changes in children’s learning-relevant affective states and to react to these changes appropriately. For this, child-robot interaction specific knowledge is necessary that could be best gathered from experts in reading and managing the affective states of young children in tutoring interactions, namely, preschool teachers.

III. EMPIRICAL BASIS

With the aim of answering the questions, which affective states occur and are important during robot-child tutoring, and how they can be detected based on the observation of a child, a qualitative approach was chosen. We used video recordings from a previous study in kindergarten and interviewed five preschool teachers on their perception and interpretation of the children’s behavior.

A. Participants

A total of five female preschool teachers were invited and interviewed as experts. They were between 36 - 61 years old ($M = 48.6$; $SD = 8.16$) and had a working experience from 16 to 42 years ($M = 29$; $SD = 8.88$).

B. Materials

With the objective of allowing the experts to observe children during robot-child tutoring in a controlled manner, video recordings from an interaction study were used. They were presented and discussed during face-to-face interviews with one interviewer. In total, video recordings of eight different children (4 female, 4 male), which varied in their level of activity and expressiveness when facing the robot, were chosen. The decision was taken to ensure that individual difference are considered in spite of the small samples. The recordings were taken in the realm of a separate study in Dutch



Fig. 1. Screenshot from one of the videos shown to the experts during the interview. The learning interaction is displayed from two perspectives.

preschools were children were tutored to learn animal names in a foreign language by means of a “I spy with my little eye...” game with a Nao robot. Here, up to four images of animals were displayed on a tablet screen, while the robot is referring to one of them using a Dutch description and the English name of the animal [31]. To choose the animal the robot mentioned, the children had to tap on the picture on the tablet. Two camera perspectives were recorded and presented to the experts to allow a frontal view on the child, but also a landscape view from the side on the whole experimental setup which includes the robot, the tablet and the child (see Fig. 1).

C. Procedure

At the beginning of each interview session, the participants were informed about the purpose and the procedure of the interview and signed an informed consent that their voice was recorded. They were instructed that they should judge the behavior and related affective state of children, which are presented in video recordings. First, a small example video was presented, which had to be commented by the experts to make sure the task was clear. Then, the interviewer started the video on a laptop and asked the expert to comment on the child’s behavior and state. After each video (one video relates to one child) the interviewer asked how the experts would react to negative changes in the child’s state, e.g., if they recognize a lack of attention, and how this could be realized with a robot. At each point in time, the interviewees were allowed to pause the video and go back to review a scene. Each expert discussed a total of four videos with the interviewer. Afterwards they were thanked for their participation and dismissed.

D. Analyses and Results

The whole interview session were recorded by means of a computer microphone, and a screen capture tool to synchronize the comments with the video recording that was played at the time. The recordings were afterwards transcribed to enable detailed content analyses of the experts’ comments. The transcripts were then analyzed regarding the following research questions:

TABLE I
CHILDREN’S STATES AND RELATED CUES

Meta-level State	State Interpretation	Behavioral Cue	n*	
Engagement	Concentration/ Thinking	eye contact	5 (4)	
		sit still	2 (2)	
		hand to head	4 (3)	
	Involvement/ Activity	mimic robots gestures	2 (2)	
		answer verbally	1 (1)	
		nodding	1 (1)	
		head-shaking	1 (1)	
	Expressive/Proud	smiling	7 (4)	
		thumb up	1 (1)	
		raise fist	1 (1)	
Disengagement	Inattentiveness/ Distraction	rub eyes	2 (1)	
		grimace	4 (4)	
		gaze away	7 (4)	
		turn away (whole body)	10(4)	
		move position (stand up, lay down)	2 (2)	
		Boredom/ Impatience	support the head with hand(s)	3 (2)
			move the head from left to right	2 (2)
	undirected finger tapping		4 (3)	
	gaze away		2 (1)	
	move position (stand up, lay down)		6 (4)	
	Negative Engagement		Skepticism	tilt head
		Disinterest	frown	1 (1)
		Averseness	lower mouth corners	1 (1)

*n is the frequency of reference to a cue; the amount of children for which the cue was observed is noted in parentheses.

- RQ1: How do experts interpret the cognitive and emotional state of children during the robot-child tutoring lessons?
- RQ2: To which behavioral cues do they refer when they remark changes (e.g., in the child’s level of attention)?
- RQ3: How would the experts react to changes in the children’s engagement from the perspective of the robot?

According to the experts descriptions of the children’s states, categories of states were derived. As listed in Table I, the childrens states can be classified into states of engagement, disengagement, and negative engagement, on a meta level (RQ1). Engagement is composed of concentration and thinking, activity and involvement, as well as expressiveness. If a child kept eye contact with the robot and tablet, and sit still, the experts interpreted their behavior as concentrated and engaged. If they mimicked the gestures the robot made, or answered verbally or nonverbally (e.g., nodding, head-shaking), they were also described as involved and thus engaged in the interaction. Likewise, expressive behaviors as smiling, or showing a thumb up were interpreted as a sign of engagement by the experts. On the other hand, behaviors that were interpreted

as signs of inattentiveness and distraction, or boredom, were regarded as indicators of disengagement. For instance, rubbing eyes, gazing away, or frequent changes of the seating position were interpreted as inattentiveness. Additionally, supporting ones head with the hands, undirected tapping with the fingers, and gazing away, were (among others, cf. Table I) named as remarkable behaviors that demonstrate boredom and disengagement. Finally, the category negative engagement contains negative states like skepticism and averseness. These states were related to frowning, lowering mouth corners, and head-tilt (RQ2).

Each interaction with the robot varied according to individual differences of the children (e.g., age, self-confidence). Hence, we counted for each behavioral cue, how many times it was mentioned by different experts for different children. If two experts observed a cue for one child as relevant it was counted as two; but if one expert mentioned one cue for one child several times it was counted as one. To reflect on the occurrence of the cues over different children, it was further listed for how many different children the cue was observed (see Table I numbers in parentheses).

The results indicate that eye contact ($n = 4$ children), smiling ($n = 4$), and self-touches to the head ($n = 3$) were interpreted as a sign of engagement for multiple children in the video recordings. Regarding disengagement, making grimaces ($n = 4$), gazing away ($n = 7$), turning away ($n = 4$), moving the position ($n = 2$), and finger tapping ($n = 3$) were observed across several children. As a sign of negative engagement, head tilt was for several children ($n = 3$) interpreted as showing skepticism. Instead, giving verbal answers, nodding, head-shake, eye rub, frowning, and lowered mouth corners were only addressed for one child, respectively, and appear hence less informative. Note that the counts refer to the spontaneous mention of the cue per child and that the cues were overall mentioned repeatedly over the course of the interaction.

Furthermore, we asked the experts how they would intervene to keep children engaged in the interaction from the robots point of view (RQ3). Their suggestions were summarized into categories of potential actions to re-engage children in the tutoring with the robot (Table II).

Parts of the experts suggestions can be regarded as preventive strategies that can be employed in the interaction from the outset. These are general strategies to keep children engaged in an interaction as allowing multi-modal interactions (here: add speech) or more expressive robot behavior (e.g., gestures, movements). Beyond that, actions were mentioned that can be useful to re-engage children in an ongoing interaction after their engagement was lowered (repair actions, see Table II). The robot could for example suggest alternative activities to get the child’s attention back (e.g., play a game). In some cases, it will even be necessary to stop the tutoring for a break according to the expert’s opinions. Moreover, it was suggested that the difficulty of the task should be increased if signs of disengagement are recognizable.

TABLE II
POSSIBLE ACTIONS MENTIONED BY THE EXPERTS

Preventive actions	Paraphrases	n*
Include verbal input	It would be more motivating for the child if it should talk to the robot (expert 2, video 2)	3
Heighten robot’s activity (e.g., move head)	The interaction would be more engaging if the robot moves. (expert 2, video 2)	3
Repair actions		
React to the child’s behavior/ give feedback	The robot should react to the behavior of the child, e.g., tell him/her to sit down again. (expert 5, video 1)	4
Change task difficulty	The task should increase in difficulty to get the child’s attention back. (expert 1, video 3)	1
Include alternative activities (e.g., play a game; stand up)	The robot could ask the child to stand up and move around, so that he/she is ready to listen again afterwards. (expert 3, video 2)	4
Allow a break	A break or a continuation at another day could be helpful to get the attention back (expert 2, video 1)	2

*n is the amount of experts out of the 5 experts that mentioned the strategy.

E. Discussion

In summary, the analyses of the expert interviews revealed that preschool teachers agree on the interpretation of several child behaviors as signs of (dis-)engagement. The behavioral cues that were identified during robot-child tutoring were changes in gaze direction (eye contact versus gaze away), body posture (turn away, stand up, lay down), or facial expressions (smiling). These cues that have been identified can be used to narrow down the feature space in affective state recognition. We note, though, that the small amount of video samples restricts the significance of our findings. However, a frequent, independent naming of the most relevant cues by different experts for different children points to the importance of these cues for detecting the affective state of children. Interestingly, the majority of these cues can be recorded by means of non-obtrusive technologies (e.g., video cameras, Microsoft Kinect) and can be extracted using existing tools (e.g., Affdex, see above). Building on this, the following section lays out a conceptual approach to interpret and respond to changes in the child’s state during robot-child tutoring interactions.

IV. AFFECTIVE STATE MANAGEMENT MODEL

A. Tracing the Affective State

The first step is to combine the different cues mentioned in Section III into higher-level states and to trace them over time. As a first approach, this can be achieved using a naive Bayesian classifier that determines the hidden internal state E that is assumed to independently cause cues C_1, C_2, \dots, C_n . Since cues need to be integrated into coherent belief updates over time, the corresponding belief must be updated every time step according to a dynamic Bayesian model $P(E^{t+1}|C_i^{t+1}, E_t)$.

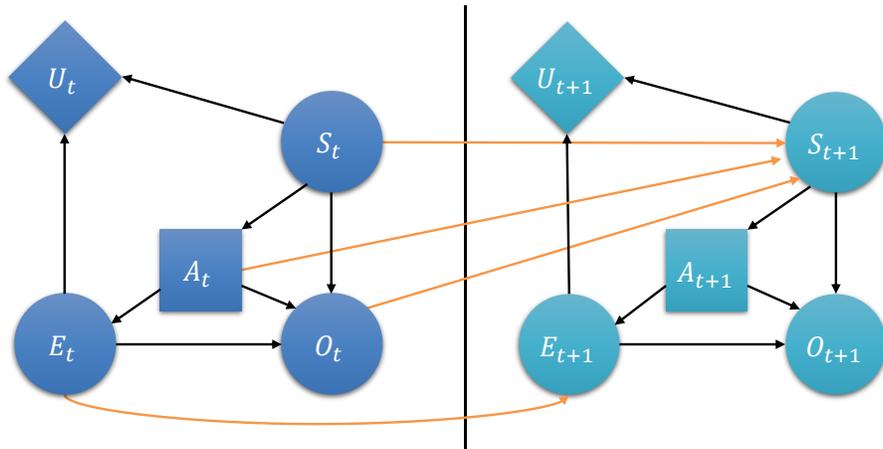


Fig. 2. Here the adaptive Bayesian Knowledge Tracing model is shown, consisting of the belief regarding the mastery of a skill S_t , the observation (response) O_t to an action A_t , the affective state E_t of the learner and the expected value U_t of a chosen chain of actions.

Variables E and C_i are directly based on the results of the expert interviews. We focus on the most reliable and explicit cues that can be tracked with current technology. Thus we base the model on those cues that were frequently mentioned for several children (cf. Table I). Since most cues from the negative engagement group were only mentioned once, and “head tilt” is difficult to track due to the danger of mixing it up with moving the head from side to side (from the disengagement group), we focus on signs of engagement and disengagement in the first stage of the model’s development. Engagement and disengagement can be regarded as opposing poles on a continuum of engagement. Hence, we combine them into the meta state variable E_t that is called *interaction engagement*. Cues that were identified as indicating engagement will have a positive effect on this state, while all cues related to disengagement will have a negative impact.

B. Managing the Affective State

After computing the belief update for interaction engagement, the next step is to determine whether and how the robot tutor should act. To this end, we include the belief variable E into our previously developed approach based on Bayesian Knowledge Tracing [2] (see Fig. 2). According to this model, the belief over the learners mastery of a certain skill S_t explains the observed answer O_t to a given teaching-task A_t selected to address this skill S_t . We add the state variable E_t as well as an utility value U_t , which represents the expected value of a chosen chain of pedagogical and affective actions. E_t is assumed to influence the students answer to a task, e.g. if the student is disengaged there may be a higher probability of observing a wrong answer as she may not have understood the task description. This information will also affect the belief update for the currently addressed skill, so that a wrong answer will have a lower impact when the student is disengaged.

Although experts’ agreed on the identification of the behavioral cues, the interpretation of these cues should be regarded carefully since one behavior could have distinct meanings depending on the situation and the specific child. For the

realization of a general model, the expert information is useful to determine which cues are relevant to look at as a starting point. A final system must, however, be able to adapt to specific variations in the child and the situation.

Next, we need to extend the action space of A_t to actions that manage the affective state, in addition to the already present actions of addressing a certain skill with a particular task. This allows evaluating and weighing both options, teaching a skill or managing the affective state of a student. Still, the main goal is to find an action (or action sequence) from which the child will learn the most. Since the model is a Dynamic Bayesian Decision Network, this evaluation can be carried out across several time steps, where each additional time step lowers the utility gained on the basis of the increase of the skill belief. Hence, the system can decide whether it is more beneficial to first raise interaction engagement, before teaching the next skill, or the other way around.

Again, we based our selection of actions to manage affective state on the results of the expert interviews (cf. Table II). We consider only the repair actions here, out of which the change of task difficulty is already implemented in the model. Three other actions remain, which could be useful to re-engage a child in the interaction: First, directly addressing the child’s behavior, e.g., urge to sit down again or ask for attention; secondly, using alternative tasks or activities to provide a more variable interaction, e.g., ask to move around or to play a game; finally, if the interaction engagement drops significantly, the robot can propose a break and the interaction can be resumed later. All of these behaviors can be immediately included in the model as well as the robot’s behavior repertoire. Note, however, that the conditional probabilities $P(E|A)$ and $P(O|A, E, S)$ need to be defined heuristically as long as sufficient interaction data is not available.

V. SUMMARY

The present paper addressed the importance of coping with a learner’s affective state during preschool child-robot tutoring. While the automatic recognition of cues seems to

be within reach with today's technology, we are still lacking a model of which affective states are most relevant in such learning interactions, how they can be recognized, and how they should be responded to by the robot tutor. To tackle this problem, expert interviews with preschool teachers have been conducted to identify children's affective states that are relevant during robot-child tutoring. The results suggest that different categories of engagement states seem to be most important, and that experts recognize and address those states in interaction. The findings from the interviews are currently used to inform the implementation of a computational model for tracing and managing the affective and cognitive state of a child learner with a robot tutor. To this end, we have laid out how to extend a previously developed knowledge-tracing and decision-making model based on a dynamic Bayesian Decision Network. The combined model will allow for finding an action policy that combines informative and affective actions of a robot tutor to manage the internal states (both, cognitive and affective) of a child learner more thoroughly, and to ensure an optimal course of learning.

ACKNOWLEDGMENT

We thank our colleagues from Tilburg University who provided the videos for the interviews. This work was supported by the Cluster of Excellence Cognitive Interaction Technology 'CITEC' (EXC 277) at Bielefeld University, funded by the German Research Foundation (DFG), and by the L2TOR (www.l2tor.eu) project, grant number: 688014, and by the BabyRobot (www.babyrobot.eu) project, grant number: 687831, both supported by the EU Horizon 2020 Program.

REFERENCES

- [1] J. Kennedy, P. Baxter, and T. Belpaeme, "The robot who tried too hard: Social behaviour of a robot tutor can negatively affect child learning," in *Proceedings of ACM/IEEE HRI 2017*. ACM, 2015, pp. 67–74.
- [2] T. Schodde, K. Bergmann, and S. Kopp, "Adaptive Robot Language Tutoring Based on Bayesian Knowledge Tracing and Predictive Decision-Making," in *Proceedings of ACM/IEEE HRI 2017*. ACM Press, 2017, pp. 128–136.
- [3] L. Vygotsky, *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press, 1978.
- [4] S. Craig, A. Graesser, J. Sullins, and B. Gholson, "Affect and learning: an exploratory look into the role of affect in learning with autotutor," *Journal of educational media*, vol. 29, no. 3, pp. 241–250, 2004.
- [5] J. Gottlieb, P.-Y. Oudeyer, M. Lopes, and A. Baranes, "Information-seeking, curiosity, and attention: computational and neural mechanisms," *Trends in cognitive sciences*, vol. 17, no. 11, pp. 585–593, 2013.
- [6] N. Thompson and T. J. McGill, "Affective tutoring systems: enhancing e-learning with the emotional awareness of a human tutor," *International Journal of Information and Communication Technology Education*, vol. 8, no. 4, pp. 75–89, 2012.
- [7] N. Schwarz, "Emotion, cognition, and decision making," *Cognition & Emotion*, vol. 14, no. 4, pp. 433–440, 2000.
- [8] B. Lehman, S. DMello, and N. Person, "The intricate dance between cognition and emotion during expert tutoring," in *Intelligent Tutoring Systems*. Springer, 2010, pp. 1–10.
- [9] S. Petrovica and M. Pudane, "Simulation of affective student-tutor interaction for affective tutoring systems: Design of knowledge structure," in *Proceedings of EET 2016*, vol. 7, 2016.
- [10] S. DMello, N. Blanchard, R. Baker, J. Ocumpaugh, and K. Brawner, "I feel your pain: A selective review of affect-sensitive instructional strategies," *Design Recommendations for Intelligent Tutoring Systems*, vol. 2, pp. 35–48, 2014.
- [11] R. A. Sottolare, J. A. DeFalco, and J. Connor, "A guide to instructional techniques, strategies and tactics to manage learner affect, engagement, and grit," *Design Recommendations for Intelligent Tutoring Systems*, vol. 2, pp. 7–33, 2014.
- [12] B. McDaniel, S. D'Mello, B. King, P. Chipman, K. Tapp, and A. Graesser, "Facial features for affective state detection in learning environments," in *Proceedings of the CogSci 2007*, vol. 29, no. 29, 2007.
- [13] L. Devillers and L. Vidrascu, "Real-life emotions detection with lexical and paralinguistic cues on human-human call center dialogs," in *Proceedings of Interspeech 2006*, 2006.
- [14] J. Kennedy, S. Lemaignan, C. Montassier, P. Lavalade, B. Irfan, F. Papadopoulos, E. Senft, and T. Belpaeme, "Child speech recognition in human-robot interaction: evaluations and recommendations," in *Proceedings of ACM/IEEE HRI 2017*. ACM, 2017, pp. 82–90.
- [15] J. H. Kahn, R. M. Tobin, A. E. Massey, and J. A. Anderson, "Measuring emotional expression with the linguistic inquiry and word count," *The American journal of psychology*, pp. 263–286, 2007.
- [16] S. D'Mello and A. Graesser, "Automatic detection of learner's affect from gross body language," *Applied Artificial Intelligence*, vol. 23, no. 2, pp. 123–150, 2009.
- [17] D. McColl and G. Nejat, "Affect detection from body language during social hri," in *21st IEEE International Symposium on Robot and Human Interactive Communication (ROMAN)*. IEEE, 2012, pp. 1013–1018.
- [18] J. Wagner, J. Kim, and E. André, "From physiological signals to emotions: Implementing and comparing selected methods for feature extraction and classification," in *Proceedings of ICME 2005*. IEEE, 2005, pp. 940–943.
- [19] O. Villon and C. Lisetti, "A user-modeling approach to build user's psycho-physiological maps of emotions using bio-sensors," in *15th IEEE International Symposium on Robot and Human Interactive Communication (ROMAN)*. IEEE, 2006, pp. 269–276.
- [20] M. H. Immordino-Yang and A. Damasio, "We feel, therefore we learn: The relevance of affective and social neuroscience to education," *Mind, brain, and education*, vol. 1, no. 1, pp. 3–10, 2007.
- [21] A. Esposito, "Affect in multimodal information," in *Affective Information Processing*. Springer, 2009, pp. 203–226.
- [22] T. Bänziger, D. Grandjean, and K. R. Scherer, "Emotion recognition from expressions in face, voice, and body: the multimodal emotion recognition test (mert)," *Emotion*, vol. 9, no. 5, p. 691, 2009.
- [23] A. Kapoor and R. W. Picard, "Multimodal affect recognition in learning environments," in *Proceedings of MM 2005*. ACM, 2005, pp. 677–682.
- [24] I. Arroyo, D. G. Cooper, W. Burleson, B. P. Woolf, K. Muldner, and R. Christopherson, "Emotion sensors go to school," in *Proceedings of AIED 2009*, vol. 200, 2009, pp. 17–24.
- [25] L. Shen, M. Wang, and R. Shen, "Affective e-Learning: Using emotional data to improve learning in pervasive learning environment related work and the pervasive e-learning platform," *Educational Technology & Society*, vol. 12, pp. 176–189, 2009.
- [26] S. Alexander, A. Sarrafzadeh, S. Hill *et al.*, "Easy with eve: A functional affective tutoring system," in *Workshop on Motivational and Affective Issues in ITS*. Citeseer, 2006, pp. 5–12.
- [27] S. D'mello and A. Graesser, "Autotutor and affective autotutor: Learning by talking with cognitively and emotionally intelligent computers that talk back," *Interactive Intelligent Systems*, vol. 2, no. 4, p. 23, 2012.
- [28] G. Gordon, S. Spaulding, J. K. Westlund, J. J. Lee, L. Plummer, M. Martinez, M. Das, and C. Breazeal, "Affective personalization of a social robot tutor for children's second language skills," in *Proceedings of 30th AAAI Conference on Artificial Intelligence*. AAAI Press, 2016, pp. 3951–3957.
- [29] D. McDuff, A. Mahmoud, M. Mavadati, M. Amr, J. Turcot, and R. e. Kaliouby, "Affdex sdk: a cross-platform real-time multi-face expression recognition toolkit," in *Proceedings of CHI 2016*. ACM, 2016, pp. 3723–3726.
- [30] S. Alexander, A. Sarrafzadeh, and S. Hill, "Foundation of an affective tutoring system: Learning how human tutors adapt to student emotion," *International journal of intelligent systems technologies and applications*, vol. 4, no. 3-4, pp. 355–367, 2008.
- [31] J. de Wit, T. Schodde, B. Willemsen, K. Bergmann, M. de Haas, S. Kopp, E. Kraemer, and P. Vogt, "Exploring the Effect of Gestures and Adaptive Tutoring on Childrens Comprehension of L2 Vocabularies," in *Proceedings of the Workshop R4L at ACM/IEEE HRI 2017*, 2017.