Pay Attention! Designing Adaptive Agents that Monitor and Improve User Engagement

Conference Paper · May 2012

DOI: 10.1145/2207676.2207679

2 authors:

Daniel Szafir
University of Colorado Boulder

Bilge Mutlu
University of Wisconsin–Madison

Pay Attention! Designing Adaptive Agents that Monitor and Improve User Engagement

Dan Szafir, Bilge Mutlu
Department of Computer Sciences, University of Wisconsin–Madison
1210 West Dayton Street, Madison, WI 53706, USA
{dszafir,bilge}@cs.wisc.edu

ABSTRACT
Embodied agents hold great promise as educational assistants, exercise coaches, and team members in collaborative work. These roles require agents to closely monitor the behavioral, emotional, and mental states of their users and provide appropriate, effective responses. Educational agents, for example, will have to monitor student attention and seek to improve it when student engagement decreases. In this paper, we draw on techniques from brain-computer interfaces (BCI) and knowledge from educational psychology to design *adaptive agents* that monitor student attention in real time using measurements from electroencephalography (EEG) and recapture diminishing attention levels using verbal and nonverbal cues. An experimental evaluation of our approach showed that an adaptive robotic agent employing behavioral techniques to regain attention during drops in engagement improved student recall abilities 43% over the baseline regardless of student gender and significantly improved female motivation and rapport. Our findings offer guidelines for developing effective adaptive agents, particularly for educational settings.

Author Keywords
Educational agents; adaptive agents; human-robot interaction; immediacy; passive brain-computer interfaces (BCI); electroencephalography (EEG)

ACM Classification Keywords
H.5.2 Information Interfaces and Presentation: User Interfaces – *input devices and strategies, evaluation/methodology, user-centered design*; H.1.2 Models and Principles: User/Machine Systems – *human factors, software psychology*

INTRODUCTION
Physically and virtually embodied agents offer great potential due to their capacity to afford interaction using the full range of human communicative behavior. To know when to best utilize these behaviors, these agents must be able to perceive subtle shifts in users' emotional and mental states. For instance, one factor behind the success of the best classroom teachers and one-on-one tutors is that they improve student learning by closely following student engagement and employing strategies that seek to increase attention, motivation, and involvement in the instruction. These behaviors involve a number of *verbal* acts such as directing questions, using humor, and addressing students with their first names and *nonverbal* cues such as expressions, gestures, eye-contact, smiling, posture, and proxemity [19, 20, 39]. Researchers have shown that students with teachers who effectively employ these strategies show more interest in the subject matter and display better cognitive performance [10].

CHI'12, May 5–10, 2012, Austin, Texas, USA.

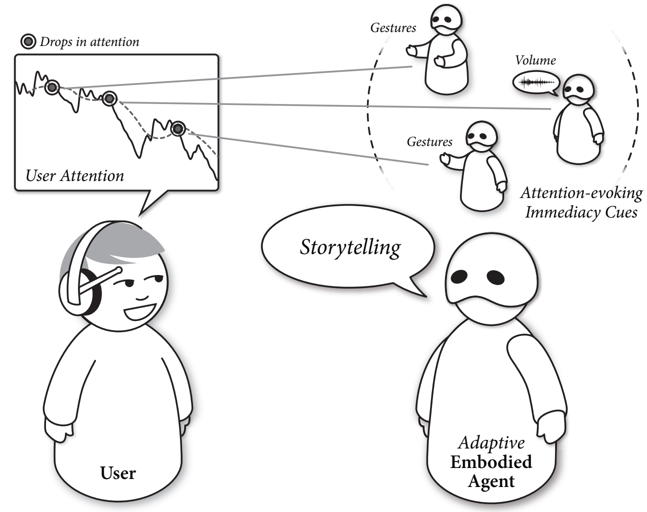

Figure 1. In our study, participants interacted with an embodied agent that monitored their attention using EEG signals in realtime and adapted its behavior to improve the discourse.

As a whole, these behaviors can be grouped under the blanket term *immediacy cues*—actions taken by speakers to decrease the psychological distance between themselves and their listeners [30]. The following excerpt from the educational literature [46] provides a vivid example of the role that these behaviors play:

Professor:	*How do you know when your teacher really means what she says?*
Third Grader:	*Well, her eyes get big and round and she* ***looks right at us****. She doesn't move and her* ***voice is a little louder****, but she talks kinda slowly. Sometimes she stands over us and looks down at us.*
Professor:	*What happens then?*
Third Grader:	*The class does what she wants!*

As in the dialogue above, human teachers change aspects of their verbal and nonverbal language to communicate to their students that they should be attending the teacher. This observation leads us to the question of how can a computer-based system monitor student engagement in a similar manner as real classroom teachers? Even if we could successfully monitor student attention, how could we then exploit that information to design agents that adapt to students and engage them in learning without disrupting the learning process? This study aims to engage these types of questions that surround the area of designing and building adaptive agents by creating a novel system in the educational domain that interacts with users and adapts in real-time to *unconscious* changes in their attentional states (Figure 1).

Educational software, from simple games such as Reading Rabbit and MathBlasters to complex and expensive language suites like Rosetta Stone, has become increasingly popular. Computer-based eduction (CBE) holds significant promise for instructional scenarios and has even shown improvement over traditional classroom education both in terms of academic achievement [1] and student motivation [44]. Further, CBE can reduce the amount of time needed for instruction as well as increase student attitudes towards learning [25]. However, there is much room for improvement as these tools "are not yet as effective as the best human tutors" [1] and "should supplement traditional instruction, not replace it" [27].

Although certain metacognitive strategies such as self-explanation can work well with CBE [1], most CBE tools lack the social and emotional aspects inherent in the traditional classroom setting. One proposed solution to this challenge is the use of virtual characters or agents, which can increase student motivation and create a more natural interaction through the use of social behaviors such as gaze, body language, and facial expressions [22]. However, modern computer systems are limited in comparison with actual teachers who can use a vast array of knowledge including their past experiences and their observations of students to inform their decisions of when and what behavior to employ.

This study seeks to answer the following question: How can we design computer-based educational tools that monitor student attention[1] and employ attention-inducing strategies to improve learning in the way human teachers do? We propose a novel approach that combines techniques and knowledge from behavioral neuroscience and educational psychology. The approach translates previous work into the physical world by utilizing a social robot that monitors real-time student attention using neural signals captured from an off-the-shelf, wireless EEG headset and models real-life educator behavior to deliver instruction. We investigate how an agent, by employing adaptive behavioral techniques to induce involvement with the instruction when significant drops in engagement are detected, might affect student achievement, motivation, and perceptions of the instructor.

[1]The terms "engagement" and "attention" are used interchangeably in this paper.

Behavioral cues	Cue affordances
Bodily cues	***Body***
Sits behind desk when teaching. (Reversed)	*Proximity*
Moves around the classroom when teaching.	*Proximity*
Sits on a desk or in a chair when teaching. (Reversed)	*Proximity*
Stands behind podium or desk when talking to the class. (Reversed)	*Proximity*
Gestures when talking to the class.	*Gestures*
Looks at the class when talking.	*Gaze*
Looks at board or notes when talking to the class. (Reversed)	*Gaze*
Has a very tense body position when talking to the class. (Reversed)	*Posture*
Has a very relaxed body position when talking to the class.	*Posture*
Smiles at individual students in the class.	*Facial Expressions*
Smiles at the class as a whole, not just individual students.	*Facial Expressions*
Touches students in the class.	*Touching*
Vocal cues	***Voice***
Uses monotone/dull voice when talking to the class. (Reversed)	*Vocal Tone*
Uses a variety of vocal expressions when talking to the class.	*Vocal Expressions*

Table 1. Verbal and nonverbal cues identified in educational settings to evaluate the effectiveness of teacher immediacy

RELATED WORK

Immediacy, Education, and Embodied Agents

Educational agents require both a means of monitoring student engagement as well as a suite of corrective actions to take when they realize that a student is no longer paying attention. Research in educational psychology has extensively investigated the various ways in which human teachers utilize vocal and nonverbal cues to achieve desired shifts in students' emotional and mental states. This literature argues that teachers display varying degrees of immediacy—the degree of perceived physical or psychological closeness between people—and that immediacy is strongly associated with the effectiveness of their teaching [23, 38].

Verbal immediacy comprises spoken content, stylistic determinations, and paralinguistic manipulations such as tone and volume, which may affect the listener positively as well as negatively [5, 21]. In general, increased immediacy can be achieved using communication content that demonstrates openness, friendliness, and empathy and an inclusive and mutual style [19].

Nonverbal immediacy includes the speaker's bodily cues and shape the listener's impressions of the speaker. In educational settings, teachers who are more expressive and use more gestures create greater immediacy and attain higher ratings by students [11]. Increasing immediacy through both verbal and nonverbal channels facilitates communication and allows people to share thoughts and feelings more easily. Furthermore, immediacy is a cyclic process whereby an increase in immediacy increases liking and brings people closer together, which, in turn, creates even greater immediacy [5]. Table 1 (adapted from [39]) lists several pivotal verbal and nonverbal behaviors that education researchers have identified in teachers who display high immediacy.

In cognitive tasks, both verbal and nonverbal immediacy have been shown to effect four major learning outcomes: student apprehension [10], motivation [11], understanding and recall of course material [23], and attitudes towards subject matter and instructors [38]. These effects have been shown to be cross-cultural [29, 28] and cross-gender [8] and can be divided into affective (student beliefs and disposition) and cognitive (recall and self-reported learning) aspects. There are two major theories as to how immediacy can be used by an instructor to positively impact student cognitive and affective learning. First, *arousal-attention theory* argues that the use of immediacy increases student arousal thus increasing student attention and engagement leading to a greater recall ability [40]. Second, *motivational theory* suggests that immediacy can increase student ambition by sparking curiosity and driving them to increased inquiry and involvement [11, 21, 40].

Education research offers strong evidence that immediacy plays a critical role in teacher-student interaction, which motivates research into how best to incorporate immediacy, presence, and proxemics in the design of educational agents. However, due to the complex nature of immediacy and the still emerging picture of how to best utilize it within the context of computer-based education, most traditional CBE systems still do not employ any form of immediacy cues. Moreover, early research into CBE has suggested that CBE systems "work better if they present themselves to students as non-human tools to assist learning rather than as emulations of human tutors" [3]. However, research in HCI provides evidence that humanlike representations improve educational outcomes such as student motivation and point to a rich design space for creating humanlike behavior such as verbal feedback that might further enhance these outcomes [33]. We argue that modern agents modeling actual human educators can surpass traditional CBE tools by utilizing arousal-attention and motivational theory.

EEG and Attention

Current computer-based tools focus on the student's comprehension of the instructional material [3] and do not consider attention and engagement, mainly due to the difficulty in inferring these cognitive states. Teachers in classroom and one-on-one tutoring situations monitor these states in real-time by reading students' behavioral cues such as direction of attention, posture, facial expressions, and responsiveness to instructional activity, and researchers have devised various self-reporting scales to measure student motivation after the fact [11, 41]. Current computational methods, however, cannot reliably capture these cues in CBE settings due to vast variability in user characteristics and environmental conditions. However, recent experiments in neuroergonomics have used electroencephalography (EEG) signals to identify subtle shifts in alertness, attention, and workload in laboratory, simulation, and real-world contexts on a second-by-second timeframe [6]. Commercially-available, low-cost wireless EEG headsets have made this approach even more feasible for brain-computer interfaces (BCIs)[34].

BCIs are a rapidly growing technology that draws on brain signals as user input. There are several forms of BCI technology currently available, which can be classified as either invasive devices that require special surgery such as electrocorticography or non-invasive methods such as EEG. Although invasive interfaces allow for better signal fidelity, non-invasive interfaces can be used by the general public, offer no risk to the user, and do not require specialized knowledge or training in their use. EEG is a particularly well-studied non-invasive method which measures event related potentials (ERPs)—positive and

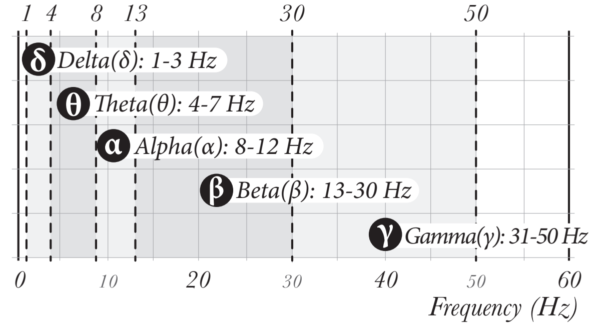

Figure 2. EEG bands and their respective frequencies.

negative charges created by inhibitory and excitatory potentials in the cerebral cortex that are created following the change in membrane conductance as neurotransmitters are released. EEG data is collected via electrodes sensitive to these changes arrayed across the scalp and is generally sequenced into signals classified by frequency (see Figure 2) that have been shown to be affected by cognitive loads in predictable ways [16]. EEG frequencies have been extensively studied and can provide insight into user mood and emotions such as anxiety, surprise, pleasure, and frustration [12, 32, 15]. Additionally, EEG measures are sensitive to cognitive states including task engagement/attention, working modality, and perception of user/machine errors [12, 13, 48].

EEG has the advantage of high temporal resolution, which offers the ability to correlate EEG data with stimuli in the external world, but has the disadvantage of low spatial resolution, which makes it difficult to determine where in the brain the signals originated from. Further, because EEG data represents a vast generalization of actual brain activity, it can be difficult to perceive small changes in user states. Finally, due to the non-invasive nature of EEG, electrode signals are highly susceptible to noise as well as influence by extraneous signals such as electromyography (EMG)—electrical signals that originate from muscles in the scalp and face as opposed to electrical signals in the brain. However, these barriers can be overcome by filtering the EEG signal and EEG has been used make accurate classifications of various cognitive tasks [26, 45]. Specifically, research has offered the following formula for calculating a signal E based on α, β, and θ waves that is highly correlated with participant task engagement [37]:

$$E = \frac{\beta}{(\alpha + \theta)} \; .$$

Although promising research is currently ongoing into how to best utilize EEG as input into controlling interfaces and robots, BCI design is impeded by the variability of user EEG waves. Traditional BCI systems utilize machine learning algorithms, which must be presented with large amounts of data in order to learn individual EEG patterns to create statistical models of user behavior [35]. These *active* systems where a user deliberately tries to control their brain activity and the system attempts to interpret and respond to the user's brain signals grew out of the goal of having systems for disabled users who could not interact with computers through standard input methods. Although research in this area has succeeded in various ways such as allowing disabled users to control a mouse cursor, these systems often require extensive tuning and training on both the part of the system and the user and are rarely generalizable across multiple users [36, 4].

Current research has shown a paradigm shift towards *passive* systems, which implicitly monitor user conscious and unconscious brain activity [47]. This implicit information is then used in conjunction with other active input by the user (e.g., vocal commands or gestures) to provide further context to the user's intended commands and state. Although research into passive BCI systems is still in its infancy, systems have already been built that can aid in the detection of user and machine errors and detect and adapt to user cognitive load [48, 18, 15]. Our research will build on these promising first steps into non-invasive, passive BCI systems as a means of improving user experience and user-system performance by creating a system that *reacts to* rather than is controlled by user EEG data. This process does not need to be trained by the user as it is based on the user's *unconscious* engagement levels. In effect, we have attempted to create an agent that passively monitors student engagement levels and reacts to them in real time in the manner that a human instructor would.

DESIGN

Our system leverages previous BCI research to create a novel system that monitors student attention in real time. When it detects significant drops in user engagement, it attempts to regain student attention through the use of immediacy cues informed by educational theory.

Detecting Drops in User Engagement

Agents who truly model human behavior must have a system that actively monitors users' attention levels to know the best time to use the verbal and nonverbal cues at their disposal. Our conjecture is that significant drops in user engagement measured from EEG might serve as close approximations of these times and that less engaging tasks should evoke more significant drops. Figure 3 shows sample engagement values from our preliminary experiments of a participant involved in tasks with varying levels of engagement. In this research,

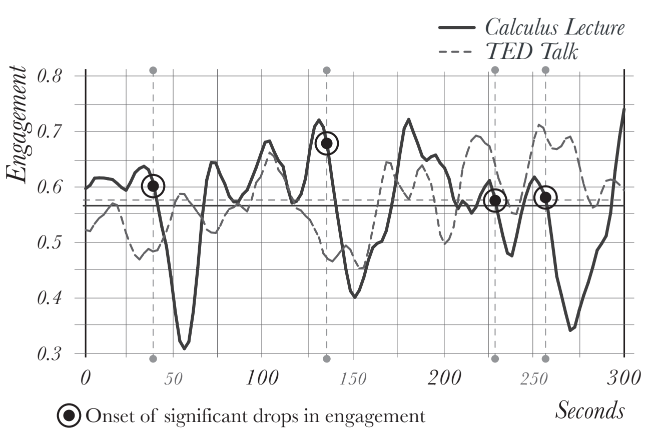

Figure 3. Engagement levels from preliminary experiments using a proprietary index of engagement as a participant watches a video of a Ted Talk (dotted line) and the same user watching a video of a calculus lecture over a six-minute period. More significant drops are seen in the calculus lecture and the onset of these drops are highlighted.

the low-cost[2] Neurosky Mindset EEG device was used as a means of testing the possibilities of off-the-shelf, low-fidelity hardware that could feasibly be used in real-world scenarios. The wireless headset gathers data using only four electrodes, which can remain dry and can quickly and easily be put on and taken off, as opposed to other EEG devices which often have many more electrodes and require special conductive liquids. The headset does not require any special training and is comfortable and usable enough for students to use in an actual classroom scenario.

The Mindset gathers EEG measurements from the FP1 region of the cortex which is known to manage learning, mental states, and concentration [17, 14, 18]. The hardware uses A1-T3 regions for grounding and filtering the signal via common mode rejection and additionally utilizes notch filters, analog and digital low- and high-pass filters, as well as proprietary algorithms to remove EMG artifacts and other noise. This filtering was verified through a pretest that assessed the presence of common artifacts such as eye-blinks. The device samples at a rate of 512Hz and is sensitive to frequencies in the range of 3-100Hz, which are broken into alpha, beta, theta, and gamma waves using Fast Fourier Transforms.

Using this device as input, we constructed a two-tiered system that analyzed EEG engagement levels and identified significant drops in attention in real-time. First, EEG levels for alpha, beta, and theta frequencies were read in from the headset. These values were smoothed using an exponentially-weighted moving average (EWMA):

$$S(t) = \begin{cases} Y(t) & : t = 1 \\ a * Y(t-1) + (1-a) * S(t-1) & : t > 1 \end{cases}$$

where S corresponds to the smoothed value produced using an EWMA, t is time, Y is the raw signal from the headset, and a is a regularization constant that is inversely proportional to the relative weighting of less recent events (a value of .015 was used in this study based on pre-test results).

Even with smoothing, the engagement signal shows high variability. To establish a consistent means of locating drops in attention levels across participants, two thresholds were created from user EEG data in real time and polled for possible losses of engagement in 15 second intervals (chosen after pretesting to identify minimum timeframe that grants sufficient data to accurately make predictions).

The first threshold compares the average slope of recent data with the average slope seen so far:

$$DT(E) = \begin{cases} 1 & : \frac{d\bar{E}(x)}{dt} < \frac{d\bar{E}(y)}{dt}, \frac{d\bar{E}(x)}{dt} < 0 \\ 0 & : otherwise \end{cases}$$

where x corresponds to a 15-second timeframe, y represents full signal seen so far, E is the smoothed engagement signal, and DT represents the derivative threshold function that outputs 1 if an immediacy cue should be displayed.

[2]At the time of this study the Neurosky Mindset was available for consumer purchase at a cost of $199.

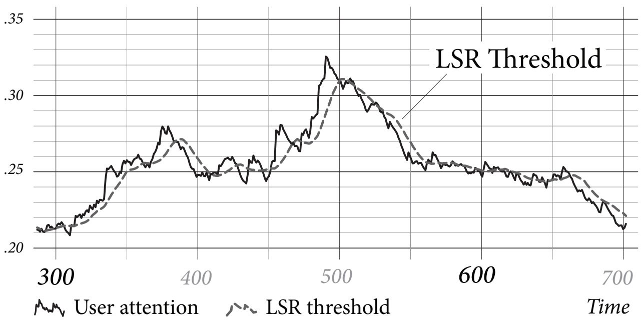

Figure 4. The LSR threshold is created by finding two lines of best fit using LSR regression. The first line is formed using all data seen so far, while the second line is based on data seen in the last time window. These values are then weighted and combined to form a rolling "average" value for user attention.

The second threshold level is based on Least-Squares Regression (LSR), which minimizes a function F, where:

$$F = \sum_{i=1}^{n} (y_i - (ax_i + b))^2 \,,$$
$$\frac{\partial F}{\partial a} = 0 \text{ and } \frac{\partial F}{\partial b} = 0 \,.$$

Here, F corresponds to the minimized function, which generates slope a and intercept b around the data represented by x and y from the start to time n. To generate the second threshold, two LSR functions are generated, one based on the past 15-second interval and one based on all the data seen so far. These values are then averaged and weighted:

$$T = .05 * F(\bar{y}) + .95 * F(\bar{x})$$

where T is the generated threshold, $F(y)$ is the function generated by LSR around all data seen so far, and $F(x)$ is the LSR function based on the data in the latest 15-second interval (see Figure 4). This creates a constantly updating "average" level for each individual user's engagement.

Our system then suggests displaying an immediacy cue by polling both the derivative threshold and the LSR threshold as long as no cue had been performed in the 15 second window immediately prior to the data window used in the calculations. This last step was put into place due to the fact that it can take time for engagement levels to rise following the instructor's use of immediacy cues. This check was enacted in response to pre-test results to avoid situations in which participants felt that the instructor gave too many cues immediately following each other.

Agent Behavior

The design of the agent's behaviors was informed by research into educator immediacy. Immediacy cues are comprised of both verbal and nonverbal behaviors that can signal accessibility or unapproachability and indicate the level of psychological distance between participants. The behaviors used by the agent in this study included modulating spoken volume and using gaze, head nodding, and gestures.

Vocal cues including tempo, pitch, tone, and volume are used by speakers to emphasize words or phrases, add emotion, and increase listener arousal [7, 43]. In our study, we increased volume when an immediacy cue was triggered due to a loss of participant attention to re-engage the student and increase instructor clarity [10]. Increased eye contact has been reported to indicate a higher level of dominance as well as affiliation and immediacy [8]. Head nodding by an instructor has also been shown to positively affect student reactions towards educators [23]. To more accurately model real-world educational scenarios, head nodding and gaze immediacy was used across all conditions to provide for a baseline interaction between the participant and the robot.

Research on gestures suggests that people use four major categories of gestures in human-human interaction: (1) iconic, (2) metaphoric, (3) deictic, (4) and beat gestures [9]. Iconic gestures are closely related to the content of speech and represent concrete events and objects, for example raising one's hands to indicate "higher." Metaphoric gestures allow the speaker to represent ethereal concepts, such as moving one's hands in a circle to show the idea of a cycle. Deictic gestures direct attention towards things directly around the speaker or to the speaker themselves, such as pointing at one's self to highlight a self-experience. Beat gestures allow the speaker to emphasize certain words and phrases and may also be linked with internal processes in speech formulation [2].

Instructor use of gestures has been shown to have a positive effect on immediacy [23] and the use of humanlike gestures by robots has been shown to facilitate interaction between humans and robots [24]. In our study, the robotic instructor utilized metaphoric, deictic, and beat gestures, to indicate that it was conveying an abstract idea, to point toward the participant and toward itself, and to add rhythmic emphasis to its speech, respectively (see Figure 5). The robot's behaviors did not include iconic gestures due to the lack of a priori knowledge of when gestures, which are triggered by a loss in participant attention, would be used during the task.

HYPOTHESES

Two hypotheses were developed for our system based on findings in BCI literature and its potential for use with adaptive educational agents:

Hypothesis 1. Instructor immediacy cues triggered by drops in EEG-monitored engagement levels will increase student attention, thereby increasing learning performance.

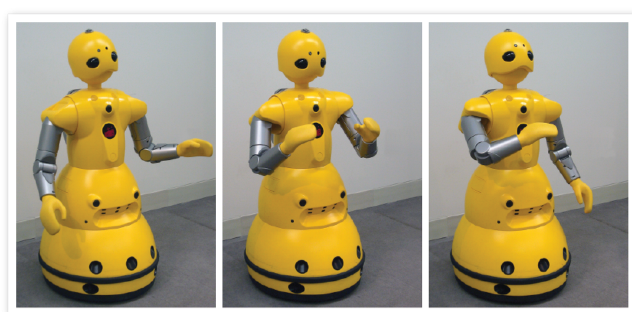

Figure 5. Examples of nonverbal immediacy gestures used by the instructor: (a) *deictic* gestures referring to the participant, (b) *metaphoric* gestures using both hands to form a gesture space containing an idea, and (c) *beat* gestures created by moving its arm rhythmically up and down.

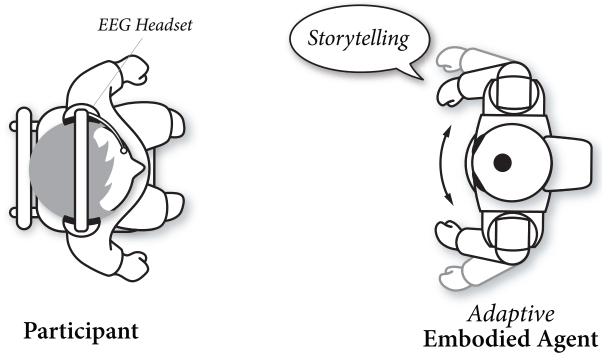

Figure 6. The spatial configuration of our experiment. In the adaptive condition, the robot used immediacy cues to regain participant attention.

Hypothesis 2. The instructor's use of immediacy cues triggered by drops in engagement will positively affect participants' motivation and evaluations of the instructor in measures of rapport.

EVALUATION

To investigate the effects of EEG-triggered adaptive immediacy cues in educational outcomes, we designed and conducted a laboratory experiment in which participants interaction with and received instruction from a humanlike robot. Below we describe the design of our experiment, our procedure and measurements, and our population.

Experimental Design

To test our hypotheses, we conducted a 3×1 between-participants study in which we manipulated the immediacy cues displayed by a Wakamaru humanlike robot, as it told participants two narrative stories. The independent variable was the introduction of the immediacy cues and included three levels: (1) low immediacy, (2) immediacy cues at random intervals, and (3) "adaptive" cues triggered by drops in their EEG-measured engagement levels. The dependent variables included participants' recall of the details of the stories, their perceptions of the robot, and self-reported learning.

Experimental Procedure

In the study, each participant was presented with a memory task that assessed the participant's recall of the details of a story narrated by the robotic teacher. After signing a consent form and being given a brief description of the experiment, participants were brought into a controlled room. Here the researcher aided the participant in putting on the wireless EEG headset and ensured good connectivity. Once the headset connection was established, the researcher left the room and the participant started interacting with the robotic instructor.

The human-robot interaction scenario consisted of five main phases: (1) introduction, (2) calibration, (3) learning, (4) distractor, and (5) evaluation, during which the robot spoke using a pre-recorded female voice modulated to a gender-neutral tone. First, the robot introduced itself and asked if the participant had any prior knowledge of the story behind the 12 signs of the Chinese Zodiac. The robot then told a pre-scripted three-minute long story about the determination of animals in the Chinese Zodiac, which was used to get baseline EEG readings that were used to build the derivative and LSR thresholds. During this calibration phase, the robot used gaze matching and head movements to make the conversation appear more natural, but did not employ other immediacy cues regardless of experimental condition. In the learning phase, the robot narrated a longer ten-minute story based on the popular Japanese folk tale "My Lord Bag of Rice." Both stories were chosen for their unfamiliarity to our participant population in order to ensure that participants had no prior task knowledge.

During the second story, instructor-participant immediacy was manipulated according to experimental condition. In the adaptive condition, the agent displayed adaptive immediacy cues by increasing its volume and employing arm gestures when a drop in engagement was identified by monitoring the participants' real-time EEG engagement data. In the random immediacy cue condition, the robot raised the volume of its voice and produced arm gestures at random intervals, the number of which was determined by the number of cues made by the instructor in the last experimental trial in the adaptive condition. In the low immediacy category, the educator told the second story in the same way it told the first story, ignoring any lapses in participant attention, although still using gaze and natural head movements that were controlled autonomously to ensure consistency between participants. While displaying an immediacy cue, the robot suspended its head movement and looked toward the participant.

After the learning phase, the instructor asked the participant four questions about the Chinese Zodiac story as a distractor task which ensured that there was a break between the learning and evaluation phases for the second story. In the last phase, the robot presented the participant with fourteen questions about the longer story to evaluate the participants' recall ability. During this question-answer period, the time between questions was controlled by the researcher behind the scenes to account for varying answer times.

Following these questions, the experimenter re-entered the room and had the participant remove the headset and fill out a post-experiment questionnaire to obtain a subjective evaluation of participant experience. Finally, participants were debriefed by the researcher and were compensated $5 for their time. The entire procedure took on average 25 minutes.

Figure 7. A participant in our study interacts with the embodied agent.

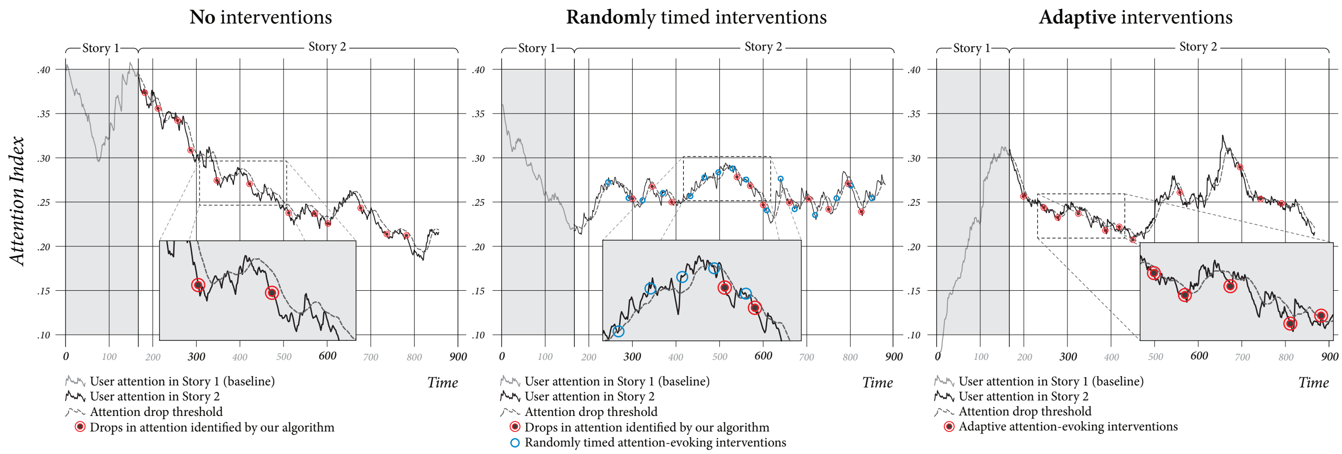

Figure 8. EEG data for three different participants in the three conditions studied. Our analysis indicates that our algorithm correctly identified real-time drops in user engagement and that instructor immediacy behavior halted downward trends in attention levels.

Participants

A total of 30 participants (15 males and 15 females) took part in this experiment. Each of the three conditions had an equal number of participants (five males and five females). All participants were native English speakers and recruited from the University of Wisconsin–Madison campus. The average age was 22.3 ($SD = 6.88$) with a range of 18-57. Prior familiarity with robots was low ($M = 3.23, SD = 1.55$) as was their familiarity with the story in the task ($M = 1.37, SD = 1.07$) in a scale of one to seven. Figure 7 shows a participant interacting with the agent.

Measurement

Objective measurements included fourteen questions that measured the participants' ability to recall the details of the "My Lord Bag of Rice" story and the participants' EEG data. Additionally, subjective measures were taken by means of a seven-point rating scale used to measure participants' responses on a post-experiment questionnaire. This questionnaire included checks on the success of our immediacy manipulations based on Richmond's Nonverbal Immediacy Scale [41] and questions regarding the participants' perception of the instructor and whether they would like to work with the robot again in the future. We utilized a modified version of Christophel's Trait Motivation Scale [11] to determine effects of the robot's behavior on student motivation.

We utilized three different checks to verify our manipulations. First, we confirmed that our algorithm successfully identified drops in attention by examining the EEG data of participants in the low immediacy condition. Second, we analyzed the EEG data of participants in the random and adaptive immediacy conditions to ensure that the agent's behaviors had a positive effect on student engagement. Finally, we constructed a five-item scale from participant responses to the post-experiment questionnaire to assess whether or not our manipulations of the robot's immediacy behavior were successful. The items asked participants how much the robot emphasized parts of story, tried to get their attention, varied the volume of its speech, used gestures, and tried to get their attention when they grew bored (Cronbach's α = .756).

RESULTS

We utilized analysis of variance (ANOVA) to analyze our data from manipulation checks and objective and subjective measurements.

Manipulation Checks – To verify that our system was working correctly, we processed the EEG data for participants in the low immediacy condition using our algorithm to identify times when the instructor would have used immediacy cues had those participants been in the adaptive condition. We then analyzed engagement levels in the 30 second timeframes before and after each possible cue using a two-way repeated measures ANOVA using participant ID as a random effect and condition, time frame, and the interaction of the two as fixed effects. Our analysis found that average engagement levels 30 seconds prior to when our algorithm would have directed the robot to re-engage the participant were significantly higher than the average engagement levels 30 seconds after this time, F(1, 658.5) = 7.54, p = .006, suggesting that our algorithm was correctly identifying losses of engagement.

Further EEG analysis yielded no significant differences in the 30-second windows before and after the robot employed behavioral strategies to regain attention in the random, F(1, 658.5) = 0.116, p = .734, and adaptive, F(1, 658.5) = 241 p = .121, conditions, showing that agent immediacy cues successfully halted downward engagement levels. Figure 8 provides a visual summary of the EEG data collected across conditions and the points of attention loss identified by our algorithm.

Our manipulation on checks on whether the participants perceived the robot to show immediacy cues showed that the manipulation was successful and that participants were able to notice when the robot used immediacy between conditions, F(2, 27) = 6.41, p = .005, η_p^2 = .322.

Objective Results – Our first hypothesis predicted that participants who received targeted immediacy cues triggered by a drop in EEG-monitored engagement levels would have better recall of the story than other participants. Our analysis confirmed this hypothesis; the number of correct answers out of fourteen questions were on average 6.30 (*SD* = 3.40), 7.44 (*SD*

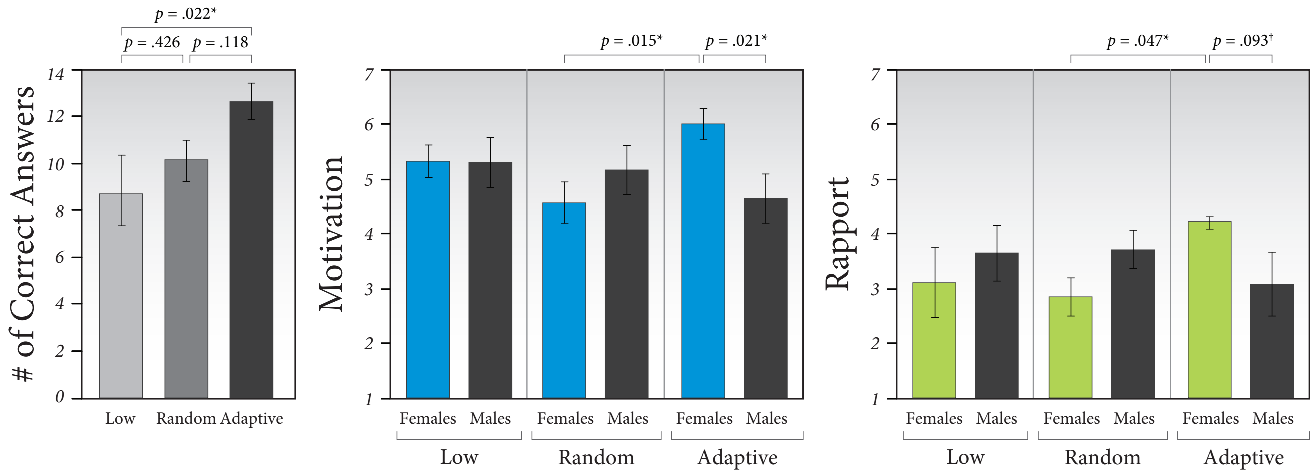

Figure 9. An analysis of our objective data supports our hypothesis that recall is positively effected by using an adaptive immediate agent.

= 1.94), and 9.00 (*SD* = 1.76) in the low immediacy, random immediacy, and adaptive immediacy conditions, respectively. These results showed that participants with an adaptive instructor outperformed the random condition by 23% and the low immediacy baseline by 43% with a significant difference between the low and adaptive immediacy levels, $F(1, 27) = 5.87$, $p = .022$, $\eta_p^2 = .177$, regardless of gender. No significant difference was found between the low and random conditions, $F(1, 27) = .652$, $p = .426$, $\eta_p^2 = .024$, or between random and adaptive conditions, $F(1, 27) = 2.60$, $p = .118$, $\eta_p^2 = .088$. A pairwise comparison, which contrasted the adaptive condition with the random and low immediacy conditions, revealed significantly improved recall accuracy in students with an adaptive instructor, $F(1, 27) = 5.43$, $p = .028$, $\eta_p^2 = .164$. Much of the variance in our model came from gender as well as users' prior familiarity with robots. When we controlled for these factors, our analysis showed even a greater difference between information recall scores in the adaptive immediacy condition and the combined scores in the low and random immediacy conditions, $F(1, 21) = 7.89$, $p = .003$, $\eta_p^2 = .291$.

Subjective Results – Our second hypothesis predicted that participants would more positively rate an adaptive agent than a random or non-adaptive one. This prediction was supported by the results from our female participants. Females felt significantly more rapport with the adaptive instructor than the random immediacy instructor, $F(1, 24) = 4.40$, $p = .047$, $\eta_p^2 = .155$, and reported that the adaptive agent motivated them significantly more $F(1, 24) = 6.83$, $p = .015$, $\eta_p^2 = .222$. Females also recognized that the adaptive instructor aided them and rated their own learning significantly higher in the adaptive case, $F(1, 24) = 7.98$, $p = .0094$, $\eta_p^2 = .244$.

Despite these promising results for females, we found no significant differences in males for rapport, motivation, or perceived learning across condition. This result could partially be explained by the fact that females reported marginally more general motivation about education and schoolwork than males did, $F(1, 24) = 3.33$, $p = .081$, $\eta_p^2 = .115$. We found no significant effects on the perceived humanlikeness of the agent or the perceived competency of the instructor in either gender. Figure 9 highlights the major results of our study.

DISCUSSION

Our first hypothesis predicted that the use of adaptive agents would significantly improve recall performance in a narrative task. Our results support this hypothesis; participants with an adaptive instructor had objectively better recall of story information. Our second hypothesis predicted that participants would react more favorably to an adaptive agent and evaluate it more positively in subjective measures. Our results partially supported this hypothesis; females felt more rapport with and motivation from an educator that adaptively displayed humanlike immediacy behaviors, while males rated all instructors equally. These results highlight the capability of utilizing EEG signals to monitor real-time user states and the benefits of designing agents that can adapt to these states.

Participants who interacted with an adaptive educational robot that displayed high immediacy had similar responses to those who interact with adaptive human educators that demonstrate high immediacy. Our results are consistent with the predictions of the arousal-attention and motivational theories; students who are taught by an instructed with high immediacy show improved recall performance, regardless of whether the instructor is human or an agent/robot. In addition, females rated the instructor with high immediacy more favorably, consistent with previous findings on gender-based differences in social perception, which indicate that females in general are better able to decode and more sensitive to nonverbal cues [42]. These results lend credence to the idea that immediacy can improve perceived learning, which can be a positive factor in student enthusiasm for classroom as well as future performance [23, 31].

Limitations & Future Work

Despite these positive results, our data showed that the instructor's use of immediacy had no significant effect on male motivation and rapport across conditions. The particular context of this study might account for this discrepancy. First, the design of the robot may have made it more difficult for males than females to connect with a "cute" robot of a small stature that used a child-like voice while instructing them. Another possibility is that a narrative task or the specific stories used in this study might affected male and female participants and their resultant experience differently. Finally, while the immediacy cues we considered in this study were limited to metaphoric, deictic, and beat gestures and volume modulation, immediacy offers a rich design space, including cues such as prosody, proximity, and facial expressions. The limited scope of our exploration prevents us from generalizing our results to situations in which the wider range of immediacy cues are used.

Regardless of their own perceptions, both males and females had an improved recall ability, which highlights the usefulness of using our immediacy cues. Moreover, the differences in educational and experiential outcomes between students who received instruction from the robotic instruction in the adaptive versus the random conditions suggest that simply displaying immediacy cues is not sufficient—immediacy cues must be strategically utilized in response to the participants' current attentional state. Although the results of this study are limited to the narrative educational setting, we feel that future work will build on our results and demonstrate the importance of adaptive agents who respond to both conscious and unconscious input from users.

While our system achieved its intended goal of improving student learning by monitoring user attention in real-time and introducing immediacy behaviors at points at which there was a significant drop in attention, open questions remain regarding our method for identifying these points. As with any measurement of neural activity, our engagement index is prone to be affected by other signals such as muscle artifacts due to the limitations of the EEG technology. Although our system employed filters to remove such artifacts, more investigation is necessary to validate that our engagement index indeed represents underlying cognitive activity free from extraneous signals. Future work should also assess the reliability of this measure across contexts, as neural signals might be affected by aspects of the social interaction and the task at hand. Additionally, we believe that the accuracy of our algorithm might be increased by exploring other means of creating tighter thresholds.

Finally, open questions remain about the interaction between agents, gender, and adaptive immediacy. Why did females show a stronger reaction to an adaptive agent? Why did males not give evaluate more positively an instructor from whom they learned objectively more? How can our system be improved to integrate EEG monitoring with other measures such as participant gaze or body posture to improve the adaptability of agents? Our future work will seek to answer these questions towards gaining a fuller understanding of how individual differences affect responses to adaptive agents and further improving our technology.

CONCLUSION

For agents to be successfully integrated into general human-computer interaction, they must be able to accurately measure and respond to conscious and unconscious user states. In this study, we created a system in which a robotic agent informed of real-time measurements of student attention obtained from EEG data employed immediacy cues that human instructors use to recapture student attention when student neural signals indicated a decline in engagement. We found significant improvements in participant recall ability when taught by an agent that adaptively employed immediacy cues compared with agents that employed these cues at random times or that showed low immediacy. Furthermore, female participants reported more motivation, rapport, and learning with the adaptive instructor. These results demonstrate that user states can be reliably evaluated using low-cost, off-the-shelf hardware and argue that the design of truly intelligent agents must incorporate the ability to react to these states in a humanlike way. We hope that this research serves as a springboard for further investigation into the field of how adaptive behavior can aid the design of effective agents to create more fluid interactions with humans.

ACKNOWLEDGMENTS

Mitsubishi Heavy Industries, Ltd., Google Inc., and the University of Wisconsin–Madison Graduate School provided support for this research. We would like to thank Jonathan Mumm, Kohei Yoshikawa, and Erin Spannan for their help in the development of our system.

REFERENCES

1. Aleven, V., and Koedinger, K. R. An effective metacognitive strategy: learning by doing and explaining with a computer-based cognitive tutor. *Cognitive Science 26*, 2 (2002), 147–179.
2. Alibali, M. W., Heath, D. C., and Myers, H. J. Effects of visibility between speaker and listener on gesture production: Some gestures are meant to be seen. *Journal of Memory and Language 44*, 2 (2001), 169–188.
3. Anderson, J. R., Corbett, A. T., Koedinger, K. R., and Pelletier, R. Cognitive tutors: Lessons learned. *The Journal of the Learning Sciences 4*, 2 (1995), 167–207.
4. Ayaz, H., Shewokis, P., Bunce, S., Schultheis, M., and Onaral, B. Assessment of cognitive neural correlates for a functional near infrared-based brain computer interface system. In *Augmented Cognition, HCII 2009*, D. Schmorrow, I. Estabrooke, and M. Grootjen, Eds., vol. 5638 of *Lecture Notes in Computer Science*. Springer Berlin / Heidelberg, 2009, 699–708.
5. Barenger, D. K., and McCroskey, J. Immediacy in the classroom: Student immediacy. *Communication Education 49* (2000), 178–186.
6. Berka, C., Levendowski, D. J., Lumicao, M. N., Yau, A., Davis, G., Zivkovic, V. T., Olmstead, R. E., Tremoulet, P. D., and Craven, P. L. EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks. *Aviation, Space, and Environmental Medicine 78*, Supplement 1 (May), B231–B244.
7. Buller, D. B., and Aune, R. K. The effects of vocalics and nonverbal sensitivity on compliance a speech accommodation theory explanation. *Human Communication Research*, 14 (1988), 301–332.
8. Burgoon, J. K., and Dillman, L. Gender, immediacy, and nonverbal communication. In *Gender, power, and communication in human relationships*, P. J. Kalbfleisch and M. J. Cody, Eds. Psychology Press, 1995.
9. Cassell, J., Steedman, M., Badler, N., Pelachaud, C., Stone, M., Douville, B., Prevost, S., and Achorn, B. Modeling the interaction between speech and gesture. In *Proc CogSci '94* (1994), 153–158.
10. Chesebro, J. L., and McCroskey, J. C. The relationship of teacher clarity and immediacy with student state receiver apprehension, affect, and cognitive learning. *Communication Education 50*, 1 (January 2001), 59–68.

11. Christophel, D. M. The relationships among teacher immediacy behaviors, student motivation, and learning. *Communication Education 39*, 4 (1990), 323–340.
12. Cutrell, E. Tan, D. BCI for passive input in HCI. In *Proc CHI '07* (2007).
13. Ferrez, P. W., and Millan. You are wrong! automatic detection of interaction errors from brain waves. In *Proc IJCAI '05* (2005).
14. Gentili, R. J., Hadavi, C., Ayaz, H., Shewokis, P. A., and Contreras-Vidal, J. L. Hemodynamic correlates of visuomotor motor adaptation by functional near infrared spectroscopy. In *Proc IEEE '10* (2010).
15. George, L., and Lecuyer, A. An overview of research on 'passive' brain-computer interfaces for implicit human-computer interaction. In *Proc ICABB '10 Workshop: "Brain-Computer Interfacing and Virtual Reality"* (2010).
16. Gevins, A., and Smith, M. E. Neurophysiological measures of cognitive workload during human-computer interaction. *Theoretical Issues in Ergonomics Science 4*, 1 (2003), 113–131.
17. Gevins, A., Smith, M. E., Leong, H., McEvoy, L., Whitfield, S., Du, R., and Rush, G. Monitoring working memory load during computer-based tasks with EEG pattern recognition methods. *Human Factors 40* (1998), 79–91.
18. Girouard, A., Solovey, E. T., and Jacob, R. J. K. Designing a passive brain computer interface using real time classification of functional near-infrared spectroscopy. *International Journal of Autonomous and Adaptive Communications Systems* (2010).
19. Gorham, J. The relationship between verbal teacher immediacy behaviors and student learning. *Communication Education 37* (1988), 40–53.
20. Gorham, J., and Christophel, D. M. The relationship of teachers' use of humor in the classroom to immediacy and student learning. *Communication Education 39* (1990), 46–62.
21. Gorham, J., and Christophel, D. M. Students' perceptions of teacher behaviors as motivating and demotivating factors in college classes. *Communication Quarterly 40*, 3 (1992), 239–252.
22. Gulz, A. Benefits of virtual characters in computer based learning environments: Claims and evidence. *Int. J. Artif. Intell. Ed. 14* (December 2004), 313–334.
23. Harris, M., and Rosenthal, R. No more teachers' dirty looks: Effects of teacher nonverbal behavior on student outcomes. applications of nonverbal communication. In *Applications of nonverbal communication*, R. E. Riggio and R. S. Feldman, Eds. Lawrence Erlbaum, 2005, 157–192.
24. Kanda, T., Ishiguro, H., Ono, T., Imai, M., and Nakatsu, R. Development and evaluation of an interactive humanoid robot "robovie". In *Proc ICRA '02*, vol. 2 (2002), 1848 –1855.
25. Kulik, C.-L. C., and Kulik, J. A. Effectiveness of computer-based instruction: An updated analysis. *Computers in Human Behavior 7*, 1-2 (1991), 75 – 94.
26. Lee, J. C., and Tan, D. S. Using a low-cost electroencephalograph for task classification in HCI research. In *Proc UIST '06* (2006), 81–90.
27. Lowe, J. Computer-based education: Is it a panacea? *Journal of Research on Technology in Education 34*, 2 (2002), 163–71.
28. McCroskey, J. C., Richmond, V. P., Sallinen, A., Fayer, J. M., and Barraclough, R. Nonverbal immediacy and cognitive learning: A cross-cultural investigation. *Communication Education 45*, 3 (July 1996), 200–211.
29. McCroskey, J. C., Richmond, V. P., Sallinen, A., Fayer, J. M., and Barraclough, R. A. A cross-cultural and multi-behavioral analysis of the relationship between nonverbal immediacy and teacher evaluation. *Communication Education 44*, 4 (October 1995), 281–291.
30. Mehrabian, A. Immediacy: An indicator of attitudes in linguistic communication. *Journal of Personality 34*, 1 (1966), 26–34.
31. Menzel, K. E., and Carrell, L. J. The impact of gender and immediacy on willingness to talk and perceived learning. *Communication Education 48*, 1 (1999), 31–40.
32. Molina, G., Tsoneva, T., and Nijholt, A. Emotional brain-computer interfaces. In *Proc ACII '09*, IEEE (2009), 1–9.
33. Mumm, J., and Mutlu, B. Human-robot proxemics: physical and psychological distancing in human-robot interaction. In *Proc HRI '11* (2011), 331–338.
34. Nijholt, A., Bos, D. P.-O., and Reuderink, B. Turning shortcomings into challenges: Brain-computer interfaces for games. *Entertainment Computing 1*, 2 (2009), 85–94.
35. Nijholt, A., and Tan, D. Brain-computer interfacing for intelligent systems. *IEEE Intelligent Systems 23*, 3 (May-June 2008), 72–79.
36. Pfurtscheller, G., Neuper, C., Guger, C., Harkam, W., Ramoser, H., Schlogl, A., Obermaier, B., and Pregenzer, M. Current trends in Graz Brain-Computer Interface (BCI) research. *IEEE Transactions on Rehabilitation Engineering 8*, 2 (2000), 216–219.
37. Pope, A. T., Bogart, E. H., and Bartolome, D. S. Biocybernetic system evaluates indices of operator engagement in automated task. *Biological Psychology 40*, 1-2 (1995), 187–195.
38. Richmond, V. P. Teacher nonverbal immediacy use and outcomes. In *Communication for Teachers*, J. Chesebro and J. McCroskey, Eds. Allyn & Bacon., 2001, 65–82.
39. Richmond, V. P., Gorham, J. S., and McCroskey, J. C. The relationship between selected immediacy behaviors and cognitive learning. *Communication Yearbook*, 10 (1987), 574–590.
40. Richmond, V. P., and McCroskey, J. C. Influencing teacher influence through immediacy. In *Power in the classroom: communication, control, and concern*, V. P. Richmond and J. C. McCroskey, Eds. Psychology Press, 1992, 101–119.
41. Richmond, V. P., McCroskey, J. C., and Johnson, A. D. Development of nonverbal immediacy scale (nis): Measures of self-and other-perceived nonverbal immediacy. *Communication Quarterly 51*, 4 (2003), 504–517.
42. Rosip, J. C., and Hall, J. A. Knowledge of nonverbal cues, gender, and nonverbal decoding accuracy. *Journal of Nonverbal Behavior 28* (2004), 267–286.
43. Scherer, K., Ladd, D., and Silverman, K. Vocal cues to speaker affect: Testing two models. *Journal of the Acoustical Society of America 76*, 5 (1984), 1346–1356.
44. Schofield, J. W. *Computers and Classroom Culture*. Cambridge University Press, 1995.
45. Tan, D. Brain-computer interfaces: applying our minds to human-computer interaction. In *Proc CHI Workshop: "What is the Next Generation of Human-Computer Interaction?"* (2006).
46. Woolfolk, A. E., and Brooks, D. The influence of teachers' nonverbal behaviors on students' perceptions and performance. *The Elementary School Journal 85* (1985), 513–528.
47. Zander, T., and Kothe, C. Towards passive brain-computer interfaces: applying brain-computer interface technology to human-machine systems in general. *Journal of Neural Engineering 8*, 2 (2011), 025005.
48. Zander, T. O., Kothe, C., Jatzev, S., and Gaertner, M. Enhancing human-computer interaction with input from active and passive brain-computer interfaces. In *Brain-Computer Interfaces*, D. S. Tan and A. Nijholt, Eds., Human-Computer Interaction Series. Springer London, 2010, 181–199.