Towards an Optimal Affect-Sensitive Instructional System of Cognitive Skills

Jacob Whitehill[1], Zewelanji Serpell[2], Aysha Foster[2],
Yi-Ching Lin[2], Brittney Pearson[2], Marian Bartlett[1], and Javier Movellan[1]
1: Machine Perception Laboratory, University of California San Diego (UCSD), La Jolla, CA, USA
2: Department of Psychology, Virginia State University (VSU), Petersburg, VA, USA
jake@mplab.ucsd.edu, zserpell@vsu.edu, aysha.foster@gmail.com, linyichen670507@yahoo.com,
pearsonbrittney@ymail.com, marni@salk.edu, movellan@mplab.ucsd.edu

Abstract

While great strides have been made in computer vision toward automatically recognizing human affective states, much less is known about how to utilize these state estimates in intelligent systems. For the case of intelligent tutoring systems (ITS) in particular, there is yet no consensus whether responsiveness to students' affect will result in more effective teaching systems. Even if the benefits of affect recognition were well established, there is yet no obvious path for creating an affect-sensitive automated tutor. In this paper we present the first steps of the OASIS project, whose goal is to develop Optimal Affect-Sensitive Instructional Systems. We present results of a pilot study to develop affect-sensitive tutors of "cognitive skills". The study was designed to: (1) assess the importance of affect to teaching, and also (2) collect training data with ecological validity that could later be used to develop an automated teacher. Experimental results suggest that affect-sensitivity is associated with higher learning gains. Behavioral analysis using automatic facial expression coding of recorded videos also suggests that smile may reveal embarrassment rather than achievement in learning scenarios.

1. Introduction

Skilled human teachers and tutors are capable of adjusting to such factors as what the student knows and does not know, and how far along he/she is towards completing a particular task. They are also able to sense a student's emotional, or affective state – including frustration, confusion, boredom, engagement, or even despair – and adjust their teaching actions accordingly. The goal of *intelligent tutoring systems* (ITS) research (e.g., [11, 3, 16]) is to implement the most important faculties possessed by human teachers in an automated system.

Until recently, ITS typically employed only a relatively impoverished set of sensors consisting of a keyboard and mouse, which amounts to only a few bits per second that they process from the student. In the OASIS project at UCSD and VSU, our goal is to go beyond these standard sensing devices and to harness more sophisticated, higher-bandwidth "affective sensors" such as automatic facial expression [13, 12] and body posture analysis [20] to yield an Optimal Affect-Sensitive Instructional System. The formal foundation of the project is the early theoretical work on optimal teaching (e.g., [4]) based on control theory, as well as more recent advances in machine learning, reinforcement learning, and computer vision.

While various researchers in the field of ITS have been migrating towards modeling affect in their instructional systems [19, 6, 17], there is, surprisingly, no firm consensus yet on whether affect sensitivity actually makes better automated teachers: In his keynote address [15] to the ITS'2008 conference in Montreal, Kurt VanLehn, a prominent ITS researcher who pioneered the Andes Physics Tutor [16], asserted that affective sensors such as automatic facial expression recognition systems were not useful in ITS, and efforts to utilize them for automated teaching were misguided. Indeed, it is conceivable that the explicit feedback given by the student to the teacher in the form of keystrokes, mouse clicks, and screen touches might constitute all that is needed for the teacher to teach well. On the other hand, we posit two reasons why modeling of affect may be important: (1) **State preference**: Certain affective states in the student may be more desirable than others. For example, a teacher might wish to avoid a situation in which the student becomes extremely upset while attempting to solve a problem. (2) **State disambiguation**: Consider a student who has been asked a question and who has not responded for several seconds. Is the student confused? Is he/she still thinking of the answer and just about to respond? Or has the student disengaged completely and perhaps even left the room? Without some form of affective sensors, these very different states may not be easily distinguished.

In this paper we tackle two problems: (1) For one particular domain of learning – cognitive skill training (described in Section 3) – we investigate whether affective state information is useful for *human* teachers to teach effectively. We analyze the utility of affective state in terms of learning gains as assessed by a pre-test and post-test on a spatial reasoning task. We use a Wizard-of-Oz (WOZ) paradigm to simulate the environment a student would face when interacting with an automated system. While conducting this experiment, we also (2) Collect data that could be used to train an automated cognitive skills teacher. These data consist of timestamped records of the student's actions (e.g., move cards on the screen), the teacher's commands (e.g., change task difficulty), and the student's face video. The ultimate goal of our research is to identify learning domains in which affect sensitivity is useful to human tutors, and to develop automated systems that utilize affective information the way human tutors do.

2. Related work

Although a number of affect-aware ITS have emerged in recent years, such as affect-sensitive versions of AutoTutor [6] and Wayang Outpost [19], it is still unclear how beneficial the affect sensitivity in these systems actually is. Some research has been conducted on the impact of the use of pedagogical agents on student's engagement and interest level [19], but studies on the impact of actual learning gains are scarce. The only study to our knowledge that specifically addresses this point is by Aist, et. al [2]: They augmented an automated Reading Tutor, designed to boost reading and speech skills by asking students to read various vocabulary words out loud, with emotional scaffolding using a WOZ framework. In their experiment, a human teacher (in a separate room) watching the student interact with the tutor could provide supplementary motivational audio prompts to the student, e.g., "You're doing fine." Compared with students in a control condition who received no emotional scaffolding, students in the affect-enhanced condition chose to persist in the learning task for a longer time. However, no statistically significant increase in learning gains was found. In their study, the *only* action the human teachers could execute was to issue a prompt – teachers could not, for instance, also change the task difficulty. Moreover, the study did not assess whether the tutors could have been as effective if they did not have access to the video of the student, i.e., if their prompts had been based solely on the student's accuracy on the task.

3. Cognitive skills training

In recent years there has emerged growing interest in "cognitive training" programs that are designed to hone basic skills such as working memory, attention, auditory processing, and logical reasoning. The motivation behind cognitive skills training is that if basic cognitive skills can be improved, performance in academic subjects such as mathematics and reading may also increase. In recent years cognitive training has been shown to correlate both with increased cognitive skills themselves [14] as well as increased performance in mathematics in minority students [8]. Certain cognitive training regimes have also been shown to boost fluid intelligence (Gf), with larger doses of training associated with larger increases in Gf [10].

In some cognitive skill training programs such as Learning Rx [1], cognitive training sessions are conducted 1-on-1 by a human trainer. Since employing a skilled human trainer for every pupil is expensive, it would be useful to automate the cognitive training process, while maintaining the benefits of having a human teacher.

3.1. Human training versus computer training

Learning Rx prescribes a dose of both 1-on-1 human-facilitated training, along with "homework" consisting of computer-based training of the same skills using the same games. In a study comparing the effectiveness of human-based versus computer-based learning with Learning Rx cognitive skill-building games, Hill, et. al found that human-based 1-on-1 training was more effective in terms of learning gains both on the cognitive skills tasks themselves as well as in associated mathematics performance [8]. When trying to develop an automated teaching system of cognitive skills, it is important to understand the causes of this result. We suggest three different hypotheses:

1. **Skill level hypothesis**: Human teachers are very adept at adapting their teaching to the the student's apparent skill level and explicit game actions.

2. **Affect-sensitivity hypothesis**: Human teachers can adapt to the affective state of the student and thereby teach more effectively.

3. **Mere presence hypothesis**: The mere presence of a human observer can positively influence the student's performance [7].

When creating an affect-sensitive teaching system, it is important to choose a learning domain in which affect-sensitivity is fruitful. If the reason why human tutors performed better than computer trainers in [8] was due to the skill level hypothesis alone, then clearly cognitive skill training is not the right domain. Similarly, if the mere presence of a human, or perhaps a human teacher's ability to converse freely with the student using perfect speech recognition is the deciding factor in effectiveness, then there is little hope that an automated system can match a human. If, however, affect sensitivity is important for the human

teacher, then it may also prove useful for automated systems. In the experiment we describe in the next section, we examine the three hypotheses above.

4. Experiment

We conducted an experiment to assess the importance of affect in cognitive skills training by an *automated* teacher. Since we have not yet built such a system, we *simulate* it using a WOZ paradigm. In WOZ experiments, a human operates behind a "curtain" (a wall, in our case), unbeknownst to the student, and controls the teaching software. For our experiment we developed a battery of three cognitive games that we developed for the Apple iPad:

- **Set**: Similar to the classic card game, Set consists of cards that contain shapes with multiple attributes including size, shape, color, and (for higher-difficulty levels) orientation. The goal is to make as many valid "sets" of 3 cards each during the time allotted. A set is valid if and only if, for each dimension, the values of the three cards are either all the same or all different.

- **Remember**: A series of randomly generated patterns appear for a brief moment (the duration depends on the current difficulty level) on the screen. If the current pattern is the same as the previous pattern, then the student presses the left button on the screen. If the pattern is different, he/she presses the right button. At each time step the student must both act (press a button) and remember the current card.

- **Sum**: Similar to Remember, a series of small integers is presented to the user at a variable rate dependent on the current difficulty level. If the sum of the current and previous numbers is even, then the user presses the left button; if it is odd, he/she presses the right button.

During piloting, students typically found the Set game the most challenging, and the other two tasks were perceived as more recreational and diverting. Hence, we used Set as the primary task that teachers should focus on. The other two tasks were provided as options to the teachers with which to give "breaks" to the students. However these breaks were to be taken only to the extent that they would help with the long term performance on Set. Before each training session, each student performed a 2-minute pre-test on Set, and after the training session (30 minutes) each student performed a 2-minute post-test on the same task. The performance metric during tests was the number of valid sets the student could make in the time allotted. A screenshot of Set (recorded on the iPad simulator) is shown in Figure 1. Students control by the game by touching an iPad-1. Student actions consist of dragging cards in the Set task, and pressing a Left or Right button during the Remember and

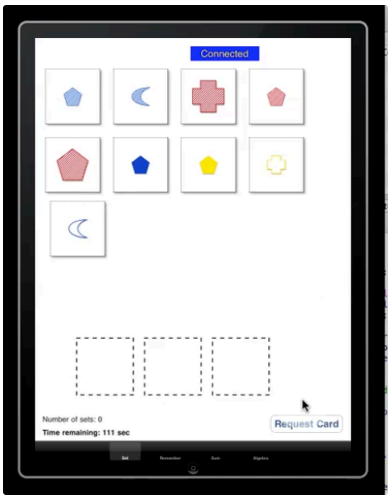

Figure 1. A screenshot of the "Set" game implemented on the Apple iPad for the cognitive skill learning experiment.

Sum tasks. The students' game inputs, along with videos of their face and upper body, were timestamped and recorded. In addition, we also recorded the teacher's actions, which consisted of increasing/decreasing task difficulty, switching tasks, giving a hint, and providing motivation in the form of pre-recorded audio prompts. The teachers were instructed to execute whatever commands they deemed necessary in order to maximize the student's learning gains on Set. These data was collected with an eye towards analyzing the teaching policies used by teachers and porting them into an automated teacher (see Section 6).

4.1. Conditions

We compared learning gains on Set across three experimental conditions:

1. **1-on-1**: The student works 1-on-1 with a human trainer who sits beside the student and makes all teaching decisions. The student is free to converse with the teacher. All of the student's and teacher's actions on the iPad, as well as a video of the student, are recorded automatically and synchronously.

2. **WOZ (full)**: The student works by him/herself on the iPad. The student is told that the iPad-based game software is controlled by an automatic teacher. In reality, it is controlled by a human trainer in another room who sees both the student's actions on the iPad as well as the student's face and upper body behavior over a videoconference. The student does not see or hear the teacher. The teacher's actions, student's actions, and student's video are all recorded automatically and synchronously.

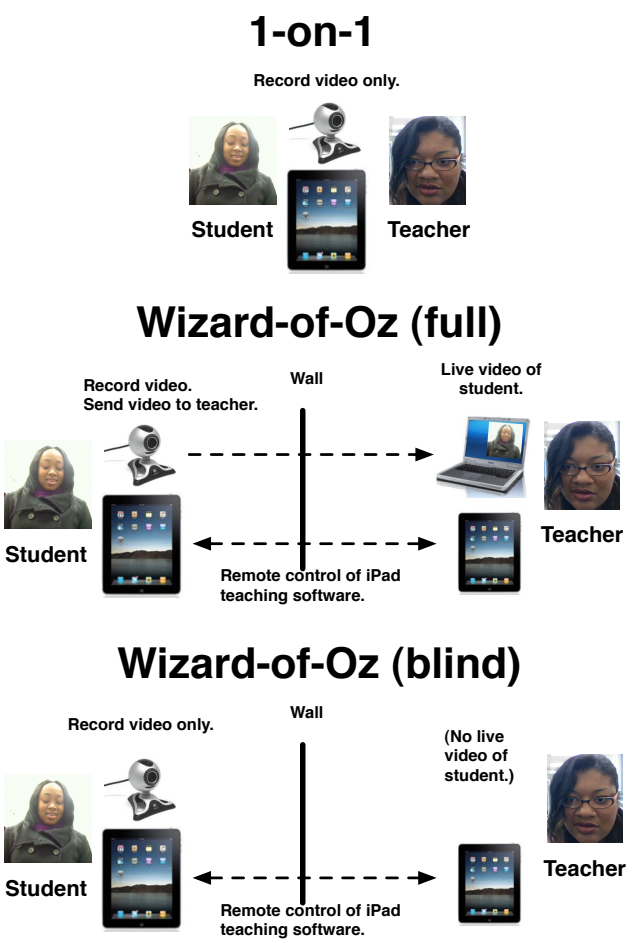

Figure 2. Three experimental conditions. **Top**: Human teacher sits with the student in a 1-on-1 training setting. **Middle**: An "automated" teacher is simulated using a Wizard-of-Oz (WOZ) technique. The iPad-based game software is controlled by a human teacher behind a wall. The teacher can see live video of the student. **Bottom**: Same as middle condition, except the teacher *cannot* see the live video of the student – the teacher sees only the student's explicit game actions.

3. **WOZ (blind)**: This condition is identical to the WOZ (full) except that the teacher *cannot* see or hear the student – the video camera records the student's behavior but does not transmit it to the teacher. In other words, the teacher is forced to teach without seeing the affective information provided by the student's face, gestures, and body posture.

Of all the students we interviewed afterwards who had participated in a WOZ condition, none suspected that the "automated teacher" was actually human.

The three conditions were designed to help distinguish which of the three hypotheses given in Section 3.1 is most valid. Consider the following possible outcomes, where performance is measured in learning gains (PostTest minus PreTest):

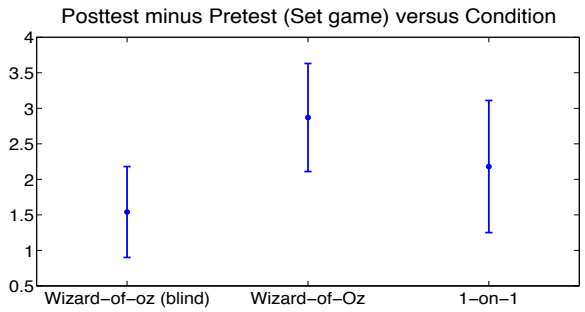

Figure 3. Average PostTest-minus-PreTest scores versus experimental condition on the "Set" spatial reasoning game. Error bars represent the standard error of the mean. In the two highest-scoring conditions (WOZ and 1-on-1) the teacher was able to observe the student's affect.

1. 1-on-1 human training is better than WOZ (full): This supports the hypothesis that merely a human's presence influences learning.

2. All three conditions are approximately equal: This supports the skill level hypothesis that affect is irrelevant to good teaching in this domain.

3. WOZ (full) is better than WOZ (blind): This supports the hypothesis that affect-sensitivity is important to effective teaching.

4. 1-on-1 is worse than the two WOZ conditions: This would suggest that a human's presence could actually detract from learning, possibly because the student felt intimidated by the human trainer's presence.

4.2. Subjects

The subject pool for this experiment consisted of 66 undergraduate students (51 female), all of whom were African-American, who were recruited from Virginia State University. Each subject was randomly assigned to one of the three conditions described above.

5. Experimental Results

5.1. Learning conditions

Performance was measured as the average PostTest minus PreTest score across each condition; results are shown in Figure 3. Although the differences (assessed by 1-way ANOVA) were not statistically significant, the two higher-performance conditions were WOZ (full) and 1-on-1. These were the two conditions in which the student's affect was visible to the teacher, thus suggesting that affect sensitivity

may indeed be important for this learning domain. Interestingly, the WOZ (full) was also higher than 1-on-1 – it is possible that the human teacher's presence was intimidating for some students and thus led to smaller learning gains.

5.2. Facial expression analysis

In addition to assessing differences in learning gains, we also examined how learning performance relates to students' facial expressions. One particular question of note is the role of smile in learning: Does occurrence of smile perhaps indicate mastery? To investigate this question we performed automatic smile detection across the videos collected of the students during the game play. We employed the Computer Expression Recognition Toolbox (CERT) [13], which is a tool for fully automatic real-time facial expression analysis from video.

To our surprise, the correlation between the average smile intensity (as estimated by CERT) over each video with PostTest-minus-PreTest performance was -0.34 ($p < 0.05$). In other words, students who learned more tended to smile *less*. This suggests that the smiles that do occur may be due more to embarrassment than to a sense of achievement. This also dovetails with findings by Hoque and Picard [9], who found that smiles frequently occur during natural episodes of frustration. Broken down by gender, the correlations were $r = -0.24$ for male ($p > 0.05$, $N = 15$), and $r = -0.35$ for female ($p < 0.05$, $N = 51$), suggesting that the effect may be more pronounced for females. We caution, however, that the reliability of CERT's smile detector on the cognitive training data has yet to be thoroughly validated against manual expression codes. Accuracy of contemporary face detection and smile detection systems on dark-skinned people in particular is known to be less reliable than for other ethnicities [18].

Examples of smiles that occurred during the experiment are shown in Figure 4. In Figure 4 (right), the subject had just made a mistake (formed an *invalid* set from 3 cards) which resulted in the game making a "buzzer" sound. Similarly, in Figure 4 (left), the teacher had just given a "giveaway" hint consisting of all 3 cards necessary to form a valid set. The student "took" the hint (made the hinted set) and then produced the expression shown, which suggests that she may have been embarrassed at needing the assistance. In contrast, the subject in Figure 5 was in the midst of scoring multiple points in rapid succession. Her facial expression during this time period shows relatively little variability in general, and no smile in particular.

6. Towards an automated affect-sensitive teaching system

The pilot experiment described above was conceived both to evaluate the hypotheses discussed in Section 3, and also to simultaneously collect training data that can be used to create an automated cognitive skills trainer. Recall that, in the WOZ (full) condition, the student interacts with an apparently "automated" iPad-based teacher, and that in this experimental condition no human was present. This interaction setting closely resembles the setting in which the student interacts with a truly automated trainer. Were training data collected from a 1-on-1 setting in which the student interacted with another human, the elicited affective states and behavior might be very different, and the collected training data might lead the automated system astray.

Figure 4. **Left**: A student who smiles as a result of receiving and acting upon a "giveaway" hint after having not scored any points for approximately 20 seconds. **Right**: A student who smiles after making a mistake, which resulted in a "buzzer" sound.

Figure 5. A student who is in the midst of scoring multiple points.

Given the "traces" of interactions between students and teachers recorded during the experiment (see Figure 6), there are several possible strategies for how to develop an affect-sensitive tutor, including rule-based expert systems, stochastic optimal control [4, 5], machine learning, or perhaps some combination of the three. In Woolf, et. al [19], for example, the authors combine manually coded rules with machine learning.

In our project we are pursuing a machine learning approach toward developing an affect-sensitive tutor:

1. Ask expert human teachers to **label the key affective states** of the student based both on the student's actions and his/her video.

2. Perform automatic facial expression recognition on the

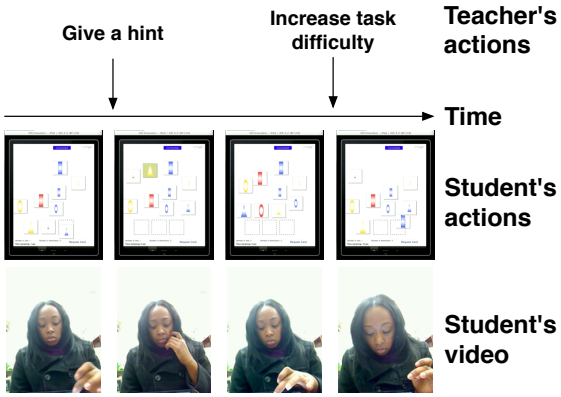

Figure 6. An example of the "traces" collected of the student's actions, the student's video, and the teacher's actions, all recorded in a synchronized manner.

student's video, in order to convert the raw video into a form more amenable to automated teaching. Classifiers such as CERT [13] output the estimated intensity of a set of facial muscle movements.

3. Train **affective state classifiers** that map from the outputs of the facial expression classifier to the higher-level states labeled in the first step.

4. Use supervised learning to compute a **policy**, i.e., a map from a history of estimated affective states extracted from the live video, the student's actions, and the teacher's previous actions, into the teacher's *next* action. The data necessary for training are available in the recorded traces.

7. Summary and further research

We have presented results from a pilot study assessing the importance of affect in automated teaching of cognitive skills. Results suggest that availability of affective state information may allow the teacher to achieve higher learning gains in the student. In addition, we have found evidence that smile during learning may indicate more embarrassment than achievement. Finally, we have proposed a methodology and software framework for collecting training data from the aforementioned experiment that can be used to train a fully automated, affect-sensitive tutoring agent. In future research we will extend the cognitive training experiment from 1 day to 6 days in an effort to elicit states with more variety, e.g., with more student fatigue.

Acknowledgement

Support for this work was provided by NSF grants SBE-0542013 and CNS-0454233, and an NSF Leadership Development Institute Fellowship from HBCU-UP to Dr. Serpell. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

[1] LearningRx, 2011. www.learningrx.com. 21

[2] G. Aist, B. Kort, R. Reilly, J. Mostow, and R. Picard. Experimentally augmenting an intelligent tutoring system with human-supplied capabilities: adding human-provided emotional scaffolding to an automated reading tutor that listens. In *Proc. Multimodal Interfaces*, 2002. 21

[3] J. R. Anderson, C. F. Boyle, and B. J. Reiser. Intelligent tutoring systems. *Science*, 228(4698):456–462, 1985. 20

[4] R. C. Atkinson. Ingredients for a theory of instruction. Technical report, Stanford Institute for Mathematical Studies in the Social Sciences, 1972. 20, 24

[5] M. Chi, K. VanLehn, and D. Litman. Do micro-level tutorial decisions matter: Applying reinforcement learning to induce pedagogical tutorial tactics. In *Intelligent Tutoring Systems*, 2010. 24

[6] S. D'Mello, R. Picard, and A. Graesser. Towards an affect-sensitive autotutor. *IEEE Intelligent Systems, Special issue on Intelligent Educational Systems*, 22(4), 2007. 20, 21

[7] B. Guerin. Mere presence effects in humans: A review. *Journal of Experimental Social Psychology*, 1986. 21

[8] O. Hill, Z. Serpell, and J. Turner. Transfer of cognitive skills training to mathematics performance in minority students. In *Poster at annual meeting of Association for Psychological Sciences*, Boston, MA, 2010. 21

[9] M. Hoque and R. Picard. Acted vs. natural frustration and delight: Many people smile in natural frustration. In *Proc. Automatic Face and Gesture Recognition (FG'11)*, 2011. 24

[10] S. Jaeggi, M. Buschkuehl, J. Jonides, and W. Perrig. Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences*, 2008. 21

[11] K. R. Koedinger and J. R. Anderson. Intelligent tutoring goes to school in the big city. *International Journal of Artificial Intelligence in Education*, 8:30–43, 1997. 20

[12] S. Koelstra, M. Pantic, and I. Patras. A dynamic texture based approach to recognition of facial actions and their temporal models. *Pattern Analysis and Machine Intelligence*, 2010. 20

[13] G. Littlewort, J. Whitehill, T. Wu, I. Fasel, M. Frank, J. Movellan, and M. Bartlett. Computer expression recognition toolbox. In *Proc. Automatic Face and Gesture Recognition (FG'11)*, 2011. 20, 24, 25

[14] R. Marachi. Statistical analysis of cognitive change with Learning Rx training procedures. Technical report, Department of Child and Adolescent Development, CSU Northridge, 2006. 21

[15] K. VanLehn. The interaction plateau: Answer-based tutoring $<$ step-based tutoring $=$ natural tutoring. In *Keynote address of the Intelligent Tutoring Systems conference*, 2008. 20

[16] K. VanLehn, C. Lynch, K. Schultz, J. Shapiro, R. Shelby, and L. Taylor. The andes physics tutoring system: Lessons learned. *International Journal of Artificial Intelligence and Education*, 15(3):147–204, 2005. 20

[17] J. Whitehill, M. Bartlett, and J. R. Movellan. Automatic facial expression recognition for intelligent tutoring systems. In *Proceedings of the CVPR 2008 Workshop on Human Communicative Behavior Analysis*, 2008. 20

[18] J. Whitehill, G. Littlewort, I. Fasel, M. Bartlett, and J. Movellan. Toward practical smile detection. *Pattern Analysis and Machine Intelligence*, 2009. 24

[19] B. Woolf, W. Burleson, I. Arroyo, T. Dragon, D. Cooper, and R. Picard. Affect-aware tutors: recognising and responding to student affect. *International Journal of Learning Technology*, 4(3):129–164, 2009. 20, 21, 24

[20] Xbox. Kinect – xbox.com, 2010. http://www.xbox.com/en-US/kinect. 20