

Analyzing Posture and Affect in Task-Oriented Tutoring

Joseph F. Grafsgaard¹, Kristy Elizabeth Boyer¹, Eric N. Wiebe², James C. Lester¹

¹Department of Computer Science, North Carolina State University, Raleigh, NC, USA

²Department of Science, Technology, Engineering, and Mathematics Education, North Carolina State University, Raleigh, NC, USA

{jgrafsg, keboyer, wiebe, lester}@ncsu.edu

Abstract

Intelligent tutoring systems research aims to produce systems that meet or exceed the effectiveness of one on one expert human tutoring. Theory and empirical study suggest that affective states of the learner must be addressed to achieve this goal. While many affective measures can be utilized, posture offers the advantages of non intrusiveness and ease of interpretation. This paper presents an accurate posture estimation algorithm applied to a computer mediated tutoring corpus of depth recordings. Analyses of posture and session level student reports of engagement and cognitive load identified significant patterns. The results indicate that disengagement and frustration may coincide with closer postural positions and more movement, while focused attention and less frustration occur with more distant, stable postural positions. It is hoped that this work will lead to intelligent tutoring systems that recognize a greater breadth of affective expression through channels of posture and gesture.

Introduction

One-on-one tutoring provided by an expert human tutor has long been recognized as a highly effective scenario for learning (Bloom, 1984). Present day research in intelligent tutoring systems (ITSs) seeks to produce learning experiences approaching or exceeding the effectiveness of expert human tutors (D'Mello and Calvo, 2011). In order to accomplish this goal, both theory (D. Wood and H. Wood, 1996; du Boulay, 2011) and empirical studies (Bloom, 1984; Cooper, Arroyo, and Woolf, 2011; D'Mello and Graesser, 2010; Lester, McQuiggan, and Sabourin, 2011; Rodrigo and Baker, 2011) suggest that affective states of the learner should be addressed.

Learners' affective experience has been fruitfully explored by analyzing numerous components of emotional expression including facial expression, prosody, physiology, and posture (Calvo and D'Mello, 2010; Zeng

et al., 2009). Posture offers a significant advantage for analysis, as it can be captured non-intrusively, unlike many physiological measures. Additionally, compared with channels such as facial expressions and prosody, posture can be analyzed at a relatively coarse level, offering substantial benefits in simplicity and interpretation.

Posture has been investigated in conjunction with various intelligent tutoring environments (Cooper et al., 2011; D'Mello and Graesser, 2010; Gonzalez-Sanchez et al., 2011). These investigations have revealed that posture is associated with many learning-centered emotions, such as boredom, confidence, confusion, delight, engagement, excitement, flow, and frustration. This relationship with learning-centered emotions highlights the importance of posture in explaining phenomena related to tutoring.

Prior studies of posture during learning have focused on pressure-sensitive seats (Cooper et al., 2011; D'Mello and Graesser, 2010; Gonzalez-Sanchez et al., 2011). While these devices are non-intrusive, they require the student to press physically onto the seat or back of the chair. Video-based posture analysis has been used to measure engagement (Sanghvi et al., 2011). However, the advent of low-cost depth sensors, such as the Microsoft Kinect, has made it possible to measure posture with far less computational effort. While sophisticated techniques exist to measure posture and movement in depth video (Glowinski et al., 2011; Shotton et al., 2011), these techniques require significant development time and model training.

In contrast, this paper presents a simple, highly accurate posture estimation technique that can be applied across single depth images recorded using a Kinect depth sensor. By leveraging regularities in the data, this algorithmic posture estimation approach requires no model training and little development time. The technique is applied to a depth video corpus of computer-mediated human-human tutoring in order to automatically identify postural shifts during learning. The automatically detected points describe learners' posture as an ordered triple, creating a data set that is subsequently compared with post-session surveys

for engagement (O'Brien and Toms, 2010) and cognitive load (Hart and Staveland, 1988). Analyses reveal that patterns of postural movement identified by the algorithm are associated with self-reported frustration, focused attention, decreased involvement, and disengagement. These findings provide new insight into the relevance of posture in the affective experience of learners during computer-mediated tutoring, and show the potential of posture estimation from depth images as a novel tool for intelligent tutoring system research.

Related Work

Prior investigations of the role of posture in learning have provided insight into the relationship between posture and affect. Studies of posture using pressure-sensitive chairs during interactions with the AutoTutor intelligent tutoring system have identified meaningful patterns with learner emotions such as boredom, confusion, delight, flow and frustration (D'Mello and Graesser, 2010). In boredom, students tended to lean backward. With delight and flow, the students leaned forward. Similarly, when experiencing confusion and frustration, students leaned forward, but at a lesser inclination than with delight and flow. In all of these affective instances, students were found to have higher arousal as expressed through greater pressure exerted on the seat.

In experiments with Wayang Outpost, a mathematics intelligent tutoring system for standardized tests, the effectiveness of posture as an indicator of affective states was explored using pressure-sensitive chairs along with facial expression, interface actions, and a pressure-sensitive mouse. In these studies, posture was found to be especially indicative of confidence and excitement (Cooper et al., 2011). Also, students in positive states were likely to be sitting in the middle of their chairs (Woolf et al., 2009).

Outside of educational contexts, there have been many efforts to understand posture and body movement. Video-based methods have been explored, such as analyzing children's posture while interacting with a game companion robot (Sanghvi et al., 2011). Notably, these techniques go beyond detecting where most of the subject's weight is being placed; they may also examine angle and curvature of a sitter's posture (Sanghvi et al., 2011), as well as movements (Glowinski et al., 2011; Sanghvi et al., 2011).

Low-cost depth sensors have enabled a new set of postural analysis techniques (Glowinski et al., 2011; Shotton et al., 2011). While techniques developed for non-depth video may also be adapted to depth recordings (Glowinski et al., 2011), approaches based on depth representations have emerged with renewed relevance (Shotton et al., 2011). In particular, a Kinect-based

approach to machine learning postural configurations across millions of examples has yielded fine-grained, robust full or cropped body segmentation (Shotton et al., 2011). However, a great investment of time and effort would be required to replicate such a model.

This paper reports on a simple posture estimation algorithm applied across Kinect depth sensor recordings collected during computer-mediated human-human tutoring. The algorithmic approach holds significant advantages in that it does not require model training, and it is highly accurate in the structured scenario of a learner interacting at a computer workstation.

Data

A computer-mediated human-human tutoring study ($N=42$) was conducted to teach computational thinking through introductory Java programming. Students and tutors interacted through a web browser interface that provided task content, basic programming functionality with the capability to compile and run programs interactively, and a textual dialogue interface. Each student was assigned one of four tutors for a series of six lessons across the semester. The lessons were structured around programming tasks that mapped to learning objectives covering a set of fundamental computational concepts. Tutors guided the task progression, and by design, previous tasks could not be revisited. Tutoring sessions were limited to a maximum length of forty minutes.

Tutorial interactions were logged to a database, with all dialogue messages, programming progress, and interface actions recorded. The Kinect depth sensors recorded depth and color images at 640x480 pixel resolution. While the Kinect sensor is capable of 30 frames per second output, we reduced our memory consumption by discarding frames for at least 100 milliseconds between each recorded frame (effectively 7-8 frames per second). Additional recordings were made for skin conductance response and facial expression, though those data are not analyzed in this paper. Figure 1 shows the research study setup at one of four student workstations. An example depth image is shown in Figure 2. The depth recordings were started and stopped manually. Due to human error, there were depth recordings for 33 of 42 subjects in Lesson 1, which constitute the data analyzed in this paper.

Study participants were selected from undergraduate students enrolled in an introduction to computing environments course for engineering students. They received partial course credit for their participation. Prior to the first tutoring interaction, students completed a pre-survey. Before each session, students completed a lesson content-based pretest. After each session, the participants answered a post-survey and posttest (identical to the

pretest). A study post-survey was also given after all tutoring sessions were completed. The focus of this paper is on the Lesson 1 tutorial interactions and their corresponding post-surveys and pre- and posttests.



Figure 1. Workstation with Kinect depth sensor, built-in webcam, skin conductance bracelet, and tutoring interface.

The post-survey items were composed of a modified User Engagement Survey (O'Brien and Toms, 2010) with the Focused Attention, Endurability, and Involvement subscales, and the NASA-TLX scale for cognitive load (Hart and Staveland, 1988) which consisted of response items for Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, and Frustration Level. The Perceived Usability, Aesthetics, and Novelty subscales of the UES were omitted as they primarily related to experience with the interface rather than the task. The UES instrument was previously validated for use with a cognitively-demanding computerized task (London et al., 2012).



Figure 2. Depth image from the Kinect depth sensor.

Posture Estimation Algorithm

A posture estimation algorithm was developed to compute posture for a given frame as a triple, (*headDepth*, *midTorsoDepth*, *lowerTorsoDepth*). An overview of the posture estimation algorithm is shown in Algorithm 1 below. The measure for *headDepth* differs from the others in that the depth value of the closest pixel in the head region is selected. This was done to account for the difference between the head, which protrudes forward, and the torso, which is often behind the desk and limbs of the student. Example output of applying the algorithm to the depth image corpus is shown in Figure 3.

Algorithm 1: POSTUREESTIMATION(*I*)

input : a depth image *I*
output : a triple of posture estimation points

- 1 *width* \leftarrow width of depth image *I*;
- 2 *height* \leftarrow height of depth image *I*;
- 3 *bottomRow* \leftarrow *height* - 1;
- 4 *center* \leftarrow *width* / 2;
- 5 *headRow* \leftarrow row of first depth pixel in center column;
- 6 *midRow* \leftarrow (*bottomRow* + *headRow*) / 2;
- 7 *lowRow* \leftarrow *midRow* + (*bottomRow* - *headRow*) / 2;
- 8 *sideBound* \leftarrow columns at \pm (5% of *width*) from *center*;
- 9 *headBound* \leftarrow rows at \pm (5% of *height*) from *headRow*;
- 10 *midBound* \leftarrow rows at \pm (5% of *height*) from *midRow*;
- 11 *lowBottom* \leftarrow *lowRow* + (*bottomRow* - *headRow*) / 4;
- 12 *lowTop* \leftarrow *lowRow* - (5% of *height*);
- 13 *headDepth* \leftarrow closest pixel in [*sideBound*, *headBound*];
- 14 *midTorsoDepth* \leftarrow farthest pixel in [*sideBound*, *midBound*];
- 15 *lowerTorsoDepth* \leftarrow farthest pixel in [*sideBound*, *lowTop* and *lowBottom*];
- 16 **return** (*headDepth*, *midTorsoDepth*, *lowerTorsoDepth*);

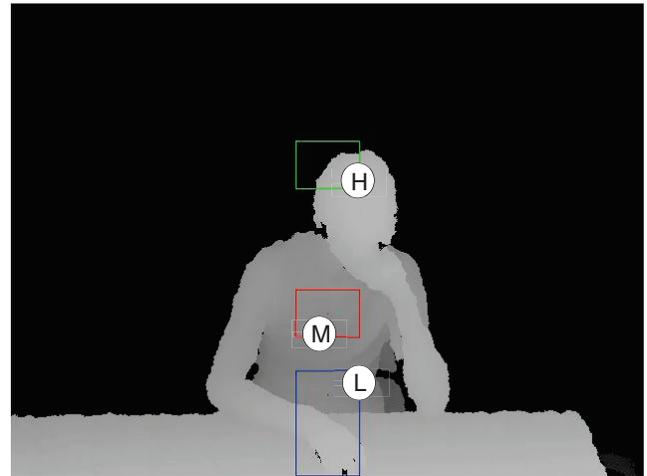


Figure 3. Output of posture estimation algorithm. Circles label detected points: (H) *headDepth*, (M) *midTorsoDepth*, (L) *lowerTorsoDepth*.

The output of the posture estimation algorithm was examined to determine how often the detected points (*headDepth*, *midTorsoDepth*, *lowerTorsoDepth*) coincided

with the head, mid torso, and lower torso/waist. Each output image was manually inspected and evaluated as correct if the points aligned to the targets. The output image was classified as erroneous if any of the points did not coincide with their target region. The accuracy was 95.1% over 1,109 depth image snapshots taken at one-minute intervals across all 33 recorded sessions for Lesson 1. The high accuracy of the algorithm is largely due to the placement of desk, computer, seat, and Kinect depth sensor. By design, the students sat in the middle of the depth recording view. Variation of the distance from the depth sensor to the student was reduced by similar positioning of chairs from the depth camera, with non-rolling chairs selected to reduce movement. Extraneous background pixels were discarded using a distance threshold. Despite these efforts, error conditions did occur when students shifted their head or torso out of frame or covered their torso or waist with their arms and hands.

Results

Once the algorithm had identified posture points, the next step was to explore whether the postural shifts correlate with students' reports of engagement and cognitive load. One frame every one minute during tutoring was labeled with its posture triple (*headDepth*, *midTorsoDepth*, *lowerTorsoDepth*). These vectors of triples were used to compute posture features for each session (Table 1). Three significant postural feature categories were identified: median distance, minimum distance, and variance of distance. The postural features of median distances indicate prevailing individual positions across each session. The postural features reflecting minimum distances are measures of how close the student moved toward the tutoring interface. The postural features of variance of distances throughout each session indicate the degree of postural movement during tutoring. Students self-reported engagement and cognitive load in post-session surveys.

Pearson correlation coefficients were computed between the postural features and self-report variables (Table 2). Median distances of *lowerTorsoDepth* (*MedLow*) and the median of all three points averaged at each one-minute interval (*MedAll*) were both negatively correlated with the frustration item from the NASA-TLX survey. That is, higher *MedLow* or *MedAll* occurs with a farther median position taken across the session. In this corpus, median and average distances were nearly identical. Thus, the farther the learners were from the tutoring interface during a majority of the session, the less the reported frustration.

The minimum of *headDepth* (*MinHead*) and the minimum of all three points averaged at each one-minute interval (*MinAll*) were both negatively correlated with the frustration item from the NASA-TLX survey. Higher

MinHead or *MinAll* indicates that a learner was farther from the sensor in the nearest position they presented across the entire session. Thus, a higher *MinHead* or *MinAll* would occur when a student leaned in less at the most extreme forward position during the session, corresponding to less reported frustration.

Table 1. Postural features produced from posture estimation algorithm points: *headDepth*, *midTorsoDepth*, and *lowerTorsoDepth*.

Postural Feature	Description
MedLow	Median value of <i>lowerTorsoDepth</i> throughout each session.
MedAll	Median value of the average of <i>headDepth</i> , <i>midTorsoDepth</i> , and <i>lowerTorsoDepth</i> at each snapshot throughout each session.
MinHead	Closest distance of <i>headDepth</i> throughout each session.
MinLow	Closest distance of <i>lowerTorsoDepth</i> throughout each session.
MinAll	Closest distance of the average of <i>headDepth</i> , <i>midTorsoDepth</i> , and <i>lowerTorsoDepth</i> at each snapshot throughout each session.
VarHead	Variance of <i>headDepth</i> distances throughout each session.
VarLow	Variance of <i>lowerTorsoDepth</i> distances throughout each session.
VarAll	Variance of the average of <i>headDepth</i> , <i>midTorsoDepth</i> , and <i>lowerTorsoDepth</i> at each snapshot throughout each session.

Table 2. Pearson correlation coefficients (*r*) between postural features and self-report variables. Only significant correlations ($p < 0.05$) are shown.

Postural Feature	Variable	<i>r</i>	<i>p</i>
MedLow	Frustration Level	-0.536	0.0013
MedAll	Frustration Level	-0.409	0.0181
MinHead	Frustration Level	-0.497	0.0032
MinLow	Focused Attention	0.362	0.0386
MinAll	Frustration Level	-0.444	0.0097
VarHead	Involvement	-0.380	0.0294
VarHead	Endurability	-0.350	0.0456
VarHead	Overall Engagement	-0.354	0.0435
VarLow	Frustration Level	0.527	0.0016
VarAll	Frustration Level	0.556	0.0008

The minimum of *lowerTorsoDepth* (*MinLow*) was positively correlated with the Focused Attention subscale of the User Engagement Survey. *MinLow* quantifies the closest that the learner's lower torso/waist approached the tutoring interface. As *MinLow* increases, the learner was farther from the interface in the closest extreme position of the lower torso/waist. The farther the student was at *MinLow*, the higher his or her reported focused attention.

The final results concern variance of the posture estimation points. This variance is similar to an aggregate *quantity of motion* (Glowinski et al., 2011) measurement. The results indicate that more head movement occurs with reports of less involvement, reduced appreciation for the tutoring session (lower score on the Endurability scale (O'Brien and Toms, 2010)), and reduced engagement overall. Greater movement of the lower torso/waist and the average of all three posture estimation points also coincide with higher reported frustration. Taken as a whole, it appears that the more the learners shifted position, the less engaged and more frustrated they were.

Discussion

The results of the correlation analyses indicate that significant relationships hold between posture and reported engagement and cognitive load factors. The median and minimum results reveal a pattern of less reported frustration as learners were farther from the tutoring interface. However, there were no significant correlations with maximum positions, so the trend does not carry to the other extreme. The results related to extremes and median of postural position are not readily comparable to past findings. For instance, in (Woolf et al., 2009) and (D'Mello and Graesser, 2010) the comparisons between students' affective states and posture were carried out at time points throughout the tutorial interaction. The results described here identify correlations between retrospective self-report measures and aggregate posture. However, the posture estimation algorithm presented here may be easily applied at fine timescale. Thus, future work will allow better comparison of the posture estimation algorithm to prior results.

Despite the difficulty of comparison discussed above, it is possible to compare postural features based on variance. Variance provides insight into postural movement throughout the session, unlike extremes or median positions. The results of the variance analyses indicated that learners who shifted position more reported less engagement and more frustration. Within the engagement subscales, students reported less involvement and reduced appreciation for the tutoring session corresponding to greater movement. Previous findings showed that learners experiencing affective states produced more movement on

their pressure-sensitive chairs (D'Mello and Graesser, 2010; Woolf et al., 2009). D'Mello and Graesser (2010) report that all affective states corresponded with an increase in movement compared to neutral. Woolf et al. (2009) report that students experiencing the least desirable cognitive-affective states of boredom or tiredness produced the greatest movement, while frustrated or angry experiences produced lesser amounts of chair movement. The present study does not include a direct measure of boredom, but the relationship between greater postural movement and experience of emotion (e.g., frustration) appears to hold in this corpus.

As shown through these correlation analyses, the posture estimation algorithm is a useful tool for investigating posture in learning. Additionally, its simplicity affords many possibilities for extension to other environments. A limitation of the posture estimation algorithm is that it does not utilize all of the information that the Kinect depth sensor produces. Additional streams of information are provided by skeletal tracking, color video, and body tracking. Skeletal tracking would be especially useful for posture estimation, as a skeletal depth model inherently calculates the angles involved in postural configurations. The Beta2 version of the Kinect SDK (Microsoft, 2011) does allow for skeletal tracking for sitting individuals, but it is not robust to occlusion. In our computer-mediated tutoring scenario, the desk covers the students' lower body, which prevents initial calibration and degrades the skeletal tracking accuracy. In testing, the skeletal tracking was not reliable for fine body movements, such as those from a seated individual. Given the algorithmic nature of the approach described here and its high accuracy within the tutoring scenario, it shows promise for use in tutoring.

Conclusion

Intelligent tutoring systems research aims to create systems that meet or exceed the effectiveness of one-on-one expert human tutoring. In order to meet this challenge, it has become apparent that both cognitive and affective factors need to be addressed. Posture provides a non-intrusive view into a learner's affective state, and has been associated with salient emotions such as boredom, confidence, confusion, delight, engagement, excitement, flow, and frustration. The introduction of low-cost depth sensors also enables posture estimation methods that are easily developed. This paper reports on a posture estimation algorithm and new evidence associating postural features with engagement and learners' affective experience. These results show that disengagement and frustration coincide with closer postural positions and more movement, while focused attention and less frustration occur with more distant, stable postural positions.

In future work, it will be important to combine the current algorithm's output with additional Kinect data streams of color images, skeletal tracking and body tracking. Additionally, more sophisticated representations of posture may be developed using the current approach as a starting point. Improvements may allow differentiation between nuanced postural configurations, such as hunched shoulders or slouching. Another direction involves recognizing finer-grained gestures from naturalistic depth video (Mahmoud et al., 2011; Mahmoud and Robinson, 2011). Algorithms may also be developed to identify gestures relevant to learning, which may have their own implications for individuals' cognitive-affective states. These lines of investigation can lead to intelligent tutoring systems that recognize a greater breadth of affective expression through channels of posture and gesture.

Acknowledgments

This work is supported in part by the North Carolina State University Department of Computer Science along with the National Science Foundation through Grants IIS-0812291, DRL-1007962 and the STARS Alliance Grant CNS-1042468. Any opinions, findings, conclusions, or recommendations expressed in this report are those of the participants, and do not necessarily represent the official views, opinions, or policy of the National Science Foundation.

References

Bloom, B. S. (1984). The 2 Sigma Problem: The Search for Methods of Group Instruction as Effective as One to One Tutoring. *Educational Researcher*, 13(6), pp. 4 16.

Calvo, R. A., and D'Mello, S. K. (2010). Affect Detection: An Interdisciplinary Review of Models, Methods, and Their Applications. *IEEE Transactions on Affective Computing*, 1(1), 18 37.

Cooper, D. G., Arroyo, I., and Woolf, B. P. (2011). Actionable Affective Processing for Automatic Tutor Interventions. In R. A. Calvo and S. K. D'Mello (Eds.), *New Perspectives on Affect and Learning Technologies* (pp. 127 140).

D'Mello, S. K., and Calvo, R. A. (2011). Significant Accomplishments, New Challenges, and New Perspectives. In R. A. Calvo and S. K. D'Mello (Eds.), *New Perspectives on Affect and Learning Technologies* (pp. 255 271).

D'Mello, S. K., and Graesser, A. C. (2010). Mining Bodily Patterns of Affective Experience During Learning. *Proceedings of the International Conference on Educational Data Mining* (pp. 31 40).

du Boulay, B. (2011). Towards a Motivationally Intelligent Pedagogy: How Should an Intelligent Tutor Respond to the Unmotivated or the Demotivated? In R. A. Calvo and S. K. D'Mello (Eds.), *New Perspectives on Affect and Learning Technologies* (pp. 41 52).

Glowinski, D., Dael, N., Camurri, A., Volpe, G., Mortillaro, M., and Scherer, K. (2011). Toward a Minimal Representation of Affective Gestures. *IEEE Transactions on Affective Computing*, 2(2), 106 118.

Gonzalez Sanchez, J., Chavez Echeagaray, M. E., Atkinson, R., and Bursleson, W. (2011). ABE: An Agent Based Software Architecture for a Multimodal Emotion Recognition Framework. *Proceedings of the Ninth Working IEEE/IFIP Conference on Software Architecture* (pp. 187 193).

Hart, S. G., and Staveland, L. E. (1988). Development of NASA TLX (Task Load Index): Results of Empirical and Theoretical Research. In P. A. Hancock and N. Meshkati (Eds.), *Human Mental Workload* (pp. 139 183).

Lester, J. C., McQuiggan, S. W., and Sabourin, J. L. (2011). Affect Recognition and Expression in Narrative Centered Learning Environments. In R. A. Calvo and S. K. D'Mello (Eds.), *New Perspectives on Affect and Learning Technologies* (pp. 85 96).

London, J., Wiebe, E., Sharek, D., Boyer, K., & Lester, J. (2012). Self Report of Engagement: Analysis of Two Scales. *To be presented at the Annual Meeting of the American Educational Research Association*.

Mahmoud, M., Baltrusaitis, T., Robinson, P., and Riek, L. D. (2011). 3D Corpus of Spontaneous Complex Mental States. *Proceedings of the International Conference on Affective Computing and Intelligent Interaction* (pp. 205 214).

Mahmoud, M., and Robinson, P. (2011). Interpreting Hand Over Face Gestures. *Proceedings of the International Conference on Affective Computing and Intelligent Interaction* (pp. 248 255).

Microsoft. (2011). Kinect for Windows SDK. Retrieved November 2011, from <http://kinectforwindows.org/>

O'Brien, H. L., and Toms, E. G. (2010). The Development and Evaluation of a Survey to Measure User Engagement. *Journal of the American Society for Information Science and Technology*, 61(1), 50 69.

Rodrigo, M. M. T., and Baker, R. S. J. d. (2011). Comparing the Incidence and Persistence of Learners' Affect During Interactions with Different Educational Software Packages. In R. A. Calvo and S. K. D'Mello (Eds.), *New Perspectives on Affect and Learning Technologies* (pp. 183 200).

Sanghvi, J., Castellano, G., Leite, I., Pereira, A., McOwan, P. W., and Paiva, A. (2011). Automatic Analysis of Affective Postures and Body Motion to Detect Engagement with a Game Companion. *Proceedings of the ACM/IEEE International Conference on Human Robot Interaction* (pp. 305 311).

Shotton, J., Fitzgibbon, A., Cook, M., Sharp, T., Finocchio, M., Moore, R., Kipman, A., et al. (2011). Real Time Human Pose Recognition in Parts from Single Depth Images. *Proceedings of the IEEE Computer Vision and Pattern Recognition Conference* (pp. 1297 1304).

Wood, D., and Wood, H. (1996). Vygotsky, Tutoring and Learning. *Oxford Review of Education*, 22(1), 5 16.

Woolf, B. P., Bursleson, W., Arroyo, I., Dragon, T., Cooper, D. G., and Picard, R. W. (2009). Affect Aware Tutors: Recognising and Responding to Student Affect. *International Journal of Learning Technology*, 4(3 4), 129 164.

Zeng, Z., Pantic, M., Roisman, G. I., and Huang, T. S. (2009). A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(1), 39 58.