Multimodal Analysis of the Implicit Affective Channel in Computer-Mediated Textual Communication

Joseph F. Grafsgaard[1], Robert M. Fulton[2], Kristy Elizabeth Boyer[1],
Eric N. Wiebe[3], James C. Lester[1]
[1]Department of Computer Science [2]Department of Psychology
[3]Department of Science, Technology, Engineering, and Mathematics Education
North Carolina State University, Raleigh, NC, USA
{jfgrafsg, rmfulton, keboyer, wiebe, lester}@ncsu.edu

ABSTRACT
Computer-mediated textual communication has become ubiquitous in recent years. Compared to face-to-face interactions, there is decreased bandwidth in affective information, yet studies show that interactions in this medium still produce rich and fulfilling affective outcomes. While overt communication (e.g., emoticons or explicit discussion of emotion) can explain some aspects of affect conveyed through textual dialogue, there may also be an underlying implicit affective channel through which participants perceive additional emotional information. To investigate this phenomenon, computer-mediated tutoring sessions were recorded with Kinect video and depth images and processed with novel tracking techniques for posture and hand-to-face gestures. Analyses demonstrated that tutors implicitly perceived students' focused attention, physical demand, and frustration. Additionally, bodily expressions of posture and gesture correlated with student cognitive-affective states that were perceived by tutors through the implicit affective channel. Finally, posture and gesture complement each other in multimodal predictive models of student cognitive-affective states, explaining greater variance than either modality alone. This approach of empirically studying the implicit affective channel may identify details of human behavior that can inform the design of future textual dialogue systems modeled on naturalistic interaction.

Categories and Subject Descriptors
I.5.4 [**Vision and Scene Understanding**]: 3D/stereo scene analysis; H.1.2 [**User/Machine Systems**]: Human factors, human information processing; J.4 [**Social and Behavioral Sciences**]: Psychology

General Terms
Algorithms, Experimentation, Human Factors, Measurement.

Keywords
Affect, computer-mediated communication, depth images, Kinect, gesture, posture, textual dialogue.

1. INTRODUCTION
Computer-mediated textual communication has become ubiquitous in recent years with pervasive use of numerous online communication channels [4]. Textual communication also plays an important role in the development of intelligent systems, mitigating challenges associated with automatic speech recognition [25] and providing a record of the interactions, which is particularly useful for tasks such as tutoring [6]. Textual communication is characterized by a limited bandwidth through which multimodal expressions of affect (e.g., facial expressions, posture, and gesture) cannot be carried. Despite this limited bandwidth, it is known that users experience a similar variety of emotional states when interacting in a textual medium [4]. Underlying the overt textual channel may be an implicit affective channel, through which participants interpret each other's cognitive-affective states [4, 14, 22].

This implicit affective channel has received little attention in affective computing, but may be the most "human" component of computer-mediated communication [21]. Understanding the implicit affective channel for computer-based systems is therefore key to providing naturalistic interaction. Achieving this goal requires empirically investigating the implicit communication of affect in a textual medium. These empirical investigations will complement affective research that focuses on the explicit affective content of textual communication [4, 10], such as work that includes emoticons, common Internet phrasing, or textual style [17, 19].

While affective content can be communicated by overt textual dialogue components, many aspects of social, cognitive, and affective information are implicitly conveyed. For example, the ways in which individuals pause or edit their messages may serve as implicit affective signals in social information processing [27]. Due in part to such affective signals, some evidence suggests that computer-mediated textual dialogue can lead to equally fulfilling interpersonal interactions as face-to-face communication [4, 22, 26]. In fact, textual communication has demonstrated utility in emotionally intensive interactions such as couples therapy [22], instant messaging chat [14], and textual communication via websites [4, 10, 26].

Although studies of the implicit affective channel in computer-mediated textual communication have identified some underlying implicit cues, nonverbal phenomena that occur beyond the computer screen have been the subject of limited study [14]. Nonverbal displays, such as facial expression [2, 15, 29], posture [2, 9, 24], and gesture [1, 12, 13] have been empirically studied in other domains and may represent an element of ground truth for cognitive-affective states. However, their roles in computer-mediated textual communication have yet to be investigated. The study of these nonverbal displays is key to empirically understanding the implicit affective channel.

ICMI'12, October 22-26, 2012, Santa Monica, California, USA.

This paper investigates the extent to which some dimensions of affect are implicitly expressed through computer-mediated textual dialogue, and further, how the perceptions of one participant correspond to the posture and gesture expressed by the other participant, even when those multimodal features were not observable through the textual dialogue channel. The focus is on two empirical questions within computer-mediated textual dialogue: 1) Does the affective interpretation by one participant coincide with reported affective states of the other participant?, and 2) Do one participant's bodily expressions coincide with reported affect of either participant?

The investigation was conducted with depth video recordings of human body position from a tutoring study carried out through computer-mediated textual dialogue. To extract posture and gesture automatically from the depth images, novel tracking techniques were designed and applied to the data set. In addition to the extracted posture and gesture data, information regarding cognitive and affective experience was collected from both tutors and students through surveys. The findings suggest that *focused attention*, *physical demand*, and *frustration* were communicated through the hypothesized implicit affective channel that accompanies textual dialogue. Moreover, postural and gestural behaviors were found to co-occur with these cognitive-affective reports, even when posture and gesture were not explicitly communicated to the other participant. These results present empirical evidence to support the notion that an implicit affective channel is at work within computer-mediated textual dialogue, and that bodily displays of posture and gesture correspond to components of this implicit affective communication. This approach yields empirical evidence of human behavior that may be used to aid in developing future naturalistic systems that engage in textual dialogue.

2. RELATED WORK

Over the past three decades, computer-mediated communication has been increasingly studied [4, 10]. Most studies of computer-mediated communication have focused on the explicit act of communication itself [4, 26]. An example of this line of investigation is a system that automatically detects emotional expression from computer-mediated textual dialogue [17]. This system analyzes affective content of messages at the level of both words and statements, and interprets emoticons and common expressions used in internet-based communication.

While studying textual communication of affect is useful, there also appears to be an implicit affective channel in computer-mediated communication. It may be necessary to investigate nonverbal affective phenomena in order to understand the human processes behind implicit affective interpretation. However, research into the relationship between nonverbal behavior and implicit interpretation of affect is scarce. A recent study examined instant messaging interactions while also recording the nonverbal behaviors unseen by the participants [14]. The participants exhibited nonverbal behaviors indicative of cognitive and affective states (postural leaning, facial expressions, gestures), even though these bodily movements were not transmitted to the recipient of the textual dialogue. A limitation of that study is that it did not use surveys or self-reports to gauge affective experience of either participant. The present study builds on that prior work by investigating posture and gesture of participants in computer-mediated textual dialogue through post-interaction surveys to gain a better understanding of specific nonverbal behaviors and participants' implicit communication of affect through a textual medium.

Both posture and gesture have been investigated in recent years for their relationship to cognitive and affective states. The relation of posture has been vigorously studied, with initial investigations utilizing pressure-sensitive chairs to identify shifting of weight [5, 9, 28]. For instance, in two studies involving intelligent tutoring systems [5, 28], boredom was associated with increased postural movement, while inconsistent postural patterns were reported across the two studies for high-arousal positive and negative affective events. In more recent posture analysis work, computer vision-based techniques have been introduced, instead of pressure-sensitive seats. In chess-playing interactions with a robot, computer vision techniques were used to identify quantity of motion, body lean angle, slouch factor, and contraction index as measurements of a child's posture as seen from a side view [24]. Quantity of motion was found to be most informative in diagnosing a child's level of engagement. In the present study, a tracking algorithm was designed to estimate posture from depth images in a frontal view, and quantity of motion was associated with cognitive-affective states such as reduced attention.

In addition to posture, gesture has long been investigated as a medium of communication [16]. Some studies have tracked gestural movements in order to visualize their co-occurrence with speech in face-to-face communication [23]. However, some gestures do not co-occur with explicit communication behavior; instead, they may coincide with cognitive and affective phenomena that occur outside the context of communication. Gesture as a cognitive-affective display was briefly touched upon in a study of nonverbal behavior during interactions with an intelligent tutoring system [28]. In that study, the gesture of a student leaning on one hand coincided with positive affective states such as joy. More recently, a broad investigation of hand-over-face gestures suggests that these gestures co-occur with cognitive-affective states such as thinking, confusion, or boredom [12, 13]. In the current study, an algorithm was developed to detect one-hand-to-face and two-hands-to-face gestures, which were found to coincide with reduced frustration or focus, respectively.

From a theoretical perspective, the functions of nonverbal expression in computer-mediated textual dialogue differ significantly from those in face-to-face interaction. Nonverbal signals in general may express *affective/attitudinal states* (what a person feels), *manipulators* (interaction with objects in the environment, including self or others), *emblems* (culture-specific signals), *illustrators* (accompanying or depicting spoken concepts) or *regulators* (signals to control flow of conversation) [21]. In textual dialogue, the bodily expressions of *emblems*, *illustrators*, and *regulators* are rare or absent [4, 10, 14, 26]. However, textual substitutes for these bodily expressions may be present (e.g., emoticons) [18]. The rarity of *emblems*, *illustrators* and *regulators* aside, nonverbal behavior of participants in computer-mediated textual dialogue contains expressions of *affective/attitudinal states* and *manipulators* [14]. In the case of the present study, bodily expressions of posture display *affective/attitudinal states* and hand-to-face gestures are *manipulators*.

3. DATA COLLECTION

The data consist of computer-mediated textual interactions in the domain of introductory computer science tutoring. Students (N=42) and tutors interacted through a web-based interface that provided learning tasks, an interface for computer programming, and textual dialogue. Each interaction was limited to forty minutes in duration. Each student interacted with a specific tutor across six

sessions on different days. Depth images (approximately 8 frames per second from a Kinect depth camera) were collected. Webcam video and skin conductance response were also recorded, but were not used in the present analyses. The student workstation configuration is shown in Figure 1 and the tutoring interface is shown in Figure 2.

Before each session, students completed a content-based pretest. After each session, students answered a post-session survey and posttest (identical to the pretest). The post-session survey items were designed to measure several aspects of engagement and cognitive load. The survey was composed of a modified User Engagement Survey [20] with Focused Attention, Endurability, and Involvement subscales, and the NASA-TLX scale for cognitive load [8], which consisted of response items for Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, and Frustration Level. Students were intentionally not asked about a wider set of emotions in order to avoid biasing their future interactions. Selected student survey items are shown in Figure 3. Additionally, tutors also reported on cognitive and affective phenomena at the end of each session. The tutor post-session survey items are shown in Figure 4.

Figure 1. Student workstation with depth camera, skin conductance bracelet, and computer with webcam

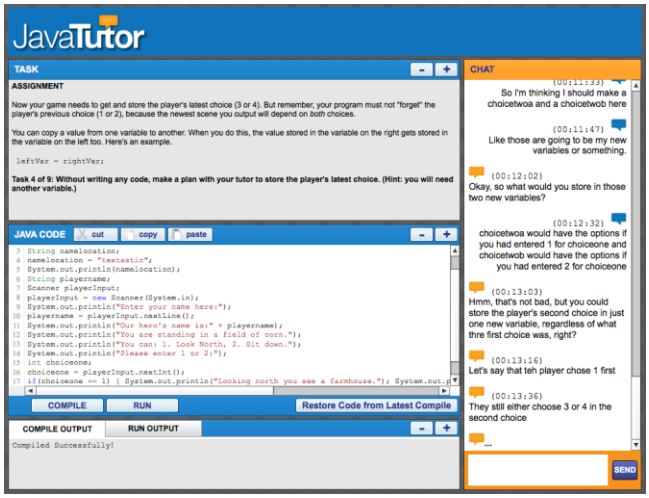

Figure 2. Screenshot of the tutoring interface

Student Post-Session Survey (from UES and NASA-TLX)

Focused Attention:
I lost myself in this learning experience.
I was so involved in my learning task that I lost track of time.
I blocked out things around me when I was working.
When I was working, I lost track of the world around me.
The time I spent working just slipped away.
I was absorbed in my task.
During this learning experience I let myself go.

Physical Demand:
How physically demanding was the task?

Frustration Level:
How insecure, discouraged, irritated, stressed, and annoyed were you?

Figure 3. Selected student post-session survey items

Tutor Post-Session Survey (reported using a 5-point scale)

Rate your agreement with the following statements:
1: Overall, the session was successful.
2: I felt like I provided cognitive support this session.
3: I helped the student finish the programming exercises more quickly than they would have on their own.
4: I helped the student master the most important concepts better than they would have on their own.
5: I was able to help the student finish the session with less effort than they would have on their own.
6: I felt like I provided emotional support this session.
7: I felt like my student was in the flow of the task.
8: I felt like my student thought the task was worthwhile.
9: I felt like my student found the task fun.
10: The student understood the computational thinking concepts.
11: The student understood the written task instructions.
12: The student understood my directions.

The student experienced the following during the lesson:
13: Anxiety (worried or uneasy about the lesson)
14: Boredom (not interested in the lesson or learning programming concepts)
15: Confusion (uncertain about some aspect of the lesson)
16: Contempt (scornful of the tutor, the lesson, or him/herself)
17: Excitement (enthusiastic or eager about the lesson)
18: Frustration (annoyed at difficulties with the tutor, the lesson, or him/herself)
19: Joy (happy about the tutor, the lesson, or him/herself)

I experienced the following during the lesson:
20-26: [tutor experience of emotion terms from items 13-19]
27: Open-ended response.

Figure 4. Tutor post-session survey of student performance and student and tutor affective experiences

4. TRACKING POSTURE AND GESTURE

In order to automatically recognize posture and gesture from the recorded interactions, tracking algorithms were designed to estimate posture and detect certain gestures from depth images. These algorithms were designed to leverage regularities in the depth recordings (e.g., student in center, frontal view). This section describes the algorithms, their output, and evaluation.

4.1 Posture Estimation

A posture estimation algorithm was designed to compute posture for a given frame as a triple, (*headDepth*, *midTorsoDepth*, *lowerTorsoDepth*), as shown in Figure 5. Prior to applying the algorithm, extraneous background pixels were discarded using a distance threshold. An overview of the posture estimation algorithm is given in Algorithm 1 below. The algorithm computes bounding regions for head, mid torso and lower torso based on the height of the top depth pixel. Then, a single point is selected from each bounding region to estimate posture. For the head, the nearest pixel is selected. For the torso points, the farthest pixel in the bounding regions is selected, as the torso was often behind the desk and arms. Distances for each posture estimation point were normalized using standard deviations from the median position for each student workstation in order to account for different camera angles. This overall approach is robust to seated postures that are occluded by a desk, a distinct advantage over the alternative of Kinect skeletal tracking.

The output of the posture estimation algorithm was evaluated manually. The performance metric was the percent of the frames in which the detected points (*headDepth*, *midTorsoDepth*, *lowerTorsoDepth*) coincided with the head, mid torso, and lower torso/waist. Two human judges individually examined images corresponding to one frame per minute of recorded video. The judges had moderate agreement on error instances with Cohen's κ=0.57. To provide a conservative measure of accuracy, the algorithm output was classified as erroneous if either judge found that any of the posture tracking points did not coincide with the target region (i.e., union of errors). Thus, the resulting accuracy was 92.4% over 1,175 depth images. Error conditions occurred primarily when students shifted their head or torso out of frame or covered their torso or waist with their arms and hands.

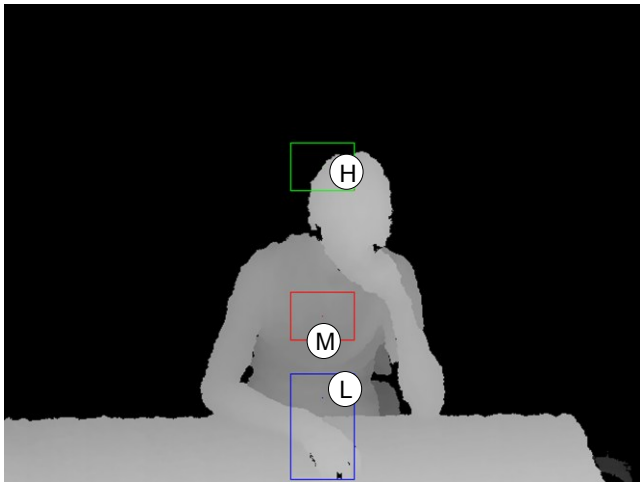

Figure 5. Detected posture points (H = *headDepth*, M = *midTorsoDepth*, L = *lowerTorsoDepth*). Bounding regions for posture point selection are also shown.

4.2 Hand-to-Face Gesture Detection

A second algorithm was developed to detect hand-to-face gestures, which have been shown to co-occur with cognitive-affective states [13]. The algorithm uses surface propagation to avoid issues of occlusion and hand deformation that pose problems for standard hand tracking techniques. Two variants of hand-to-face gestures were detected: one hand to the student's face and two hands to the student's face. Examples of detected hand-to-face gestures are shown in Figure 6. An overview of the hand-to-face gesture detection algorithm is shown in Algorithm 2.

Algorithm 1: POSTUREESTIMATION(*I*)

input : a depth image *I*
output : a triple of posture estimation points
1 *width* ⟵ width of depth image *I*;
2 *height* ⟵ height of depth image *I*;
3 *bottomRow* ⟵ *height* – 1;
4 *center* ⟵ *width* / 2;
5 *headRow* ⟵ row of first depth pixel in center column;
6 *midRow* ⟵ (*bottomRow* + *headRow*) / 2;
7 *lowRow* ⟵ *midRow* + (*bottomRow* – *headRow*) / 2;
8 *sideBound* ⟵ columns at ± (5% of *width*) from *center*;
9 *headBound* ⟵ rows at ± (5% of *height*) from *headRow*;
10 *midBound* ⟵ rows at ± (5% of *height*) from *midRow*;
11 *lowBottom* ⟵ *lowRow* + (*bottomRow* – *headRow*) / 4;
12 *lowTop* ⟵ *lowRow* – (5% of *height*);
13 *headDepth* ⟵ closest pixel in [*sideBound*, *headBound*];
14 *midTorsoDepth* ⟵ farthest pixel in [*sideBound*, *midBound*];
15 *lowerTorsoDepth* ⟵ farthest pixel in [*sideBound*, *lowTop* and *lowBottom*];
16 **return** (*headDepth*, *midTorsoDepth*, *lowerTorsoDepth*);

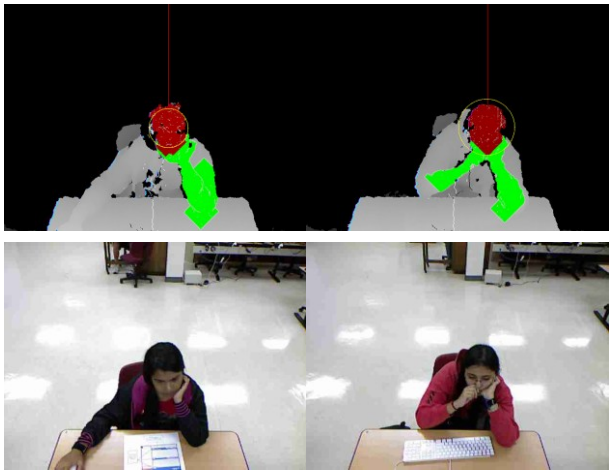

Figure 6. Detected hand-to-face gestures: one-hand-to-face (top left) and two-hands-to-face (top right). Color image frames for the detected gestures are also shown (bottom row).

The breadth-first surface propagation mentioned on line 5 of Algorithm 2 adds pixels to the set of "surface pixels" through a between-neighbors comparison of an empirically-determined surface gradient threshold. Thus, the "head surface" propagates outward from the *headPixel* (the pixel from which propagation began). If a hand-to-face gesture was detected during surface propagation (as determined by difference between mean and median distances of surface pixels from *headPixel* on line 10), then the later-propagated surface pixels were considered "hand pixels."

To evaluate the algorithm, two human judges individually examined images corresponding to one-minute snapshots of the interactions and identified one-hand-to-face and two-hands-to-face gestures. The algorithm output was compared against all instances where the judges agreed (Cohen's κ=0.96 for one-hand-to-face and κ=0.87 for two-hands-to-face). For those agreed-on instances, the accuracy of the algorithm was 92.6% across 1,170

depth images. Error cases typically involved such things as the surface propagation algorithm misidentifying clothing or hair as a hand.

Algorithm 2: HANDTOFACEGESTUREDETECTOR(*I*)

```
input  : a depth image I
output : a value indicating gesture presence or absence
headCenter ⟵ median center column selected from top 10% of rows containing non-zero depth values;
headRow ⟵ lowest row in top 10% of rows containing non-zero depth values;
headPixel ⟵ pixel location at (headCenter, headRow);
gestureDetected ⟵ false;
while performing breadth-first surface propagation do
    medianXDistance ⟵ median of horizontal distances of surface pixels from headPixel;
    meanXDistance ⟵ mean of horizontal distances of surface pixels from headPixel;
    medianYDistance ⟵ median of vertical distances of surface pixels from headPixel;
    meanYDistance ⟵ mean of vertical distances of surface pixels from headPixel;
    if |mean – median| ≥ 2.5% of mean or median do
        gestureDetected ⟵ true;
if gestureDetected do
    handPixels ⟵ surface pixels propagated after hand-to-face gesture was detected;
    if ≥33% of handPixels to upper right of headPixel do
        return NOGESTURE;
    else if ≥33% of handPixels to upper left of headPixel do
        return NOGESTURE;
    else if ≥33% of handPixels to lower left of headPixel and ≥33% of handPixels to lower right of headPixel do
        return TWOHANDSTOFACE;
    else if ≥33% of handPixels to lower left of headPixel or ≥33% of handPixels to lower right of headPixel do
        return ONEHANDTOFACE;
    else return NOGESTURE;
else return NOGESTURE;
```

5. SURVEY-BASED ANALYSES

An underlying notion of the implicit affective channel hypothesis is that tutors may have been able to identify student cognitive-affective states to some extent even when bodily movements associated with those affective states were not communicated to the tutor. To examine this, tutor perceptions of student affect were compared against student self-reports using correlational analyses in a two-step process designed to mitigate the potential for false positives (Type I error). In the first step, significant correlations were identified on a test data set drawn from the second of six tutoring sessions in which each tutor/student pair engaged (N=42). Tutor perceptions of student affect were compared against learning outcomes (posttest minus pretest) and student affect self-reports. Student affect self-reports were in turn compared against tutor reports of cognitive variables, tutor reports of student affect, and learning outcomes. Forty-three significant correlations were identified in this first step, and in the second step analyses were conducted to identify which of these significant correlations also held in a different data set, the first tutoring session, which is the main focus of the present study (N=42).[1] The two-phase analysis identified eight statistically reliable correlations, shown in Table 1.

Table 1. Significant correlations (T=tutor report; S=student report)

First Variable	Second Variable	*r*	*p*
Focused Attention^S	Helped Speed^T (Figure 4, item 3)	-0.42	0.019
	Helped Mastery^T (Figure 4, item 4)	-0.48	<0.01
	Student Confusion^T (Figure 4, item 15)	-0.39	0.029
Physical Demand^S	Student Frustration^T (Figure 4, item 18)	0.44	0.014
	Tutor Frustration^T (Figure 4, item 25)	0.42	0.019
Frustration Level^S	Student Confusion^T	0.53	<0.01
Student Confusion^T	Student Frustration^T	0.59	<0.01
Posttest Score	Student Confusion^T	-0.38	0.038

The significant correlations highlight three student cognitive-affective states: focused attention, physical demand, and frustration. Students' report of focused attention correlated with tutors' belief that they were less helpful, and tutors' belief that students were less confused. Tutors may have perceived that students who focused on the programming tasks did not need as much help to complete the tasks within the time allotted or to understand the related concepts. This result is compatible with the theory of optimal experience [3], which posits an optimally productive state of *flow* in which a student is learning well and it is often desirable not to interrupt his or her progress. Additionally, tutors may have perceived focused students as having less confusion throughout the session.

In addition to negatively correlating with students' reports of focused attention, tutor reports of student confusion were positively correlated with student self-reports of frustration. The relationship between frustration and confusion during learning may be a complex one. Specifically, frustration is typically considered to be a negative affective state, with persistent frustration referred to as a *state of stuck* [9], in which performance on the task at hand is negatively impacted. However, confusion has been theorized to be a cognitive-affective state with either positive or negative outcomes, depending on its resolution. In the theory of *cognitive disequilibrium* [7], confusion occurs with partial understanding of new knowledge, which may lead to learning when new knowledge is understood. However, it may lead to frustration when the confusion is not resolved. In the current study, positive resolutions of confusion may not have been as memorable for tutors, which could leave the tutor to report lingering confusion that may have led to student frustration. The

[1] This two-step process was used to identify correlations within one data set that generalize to another. The goal is analogous to that of statistical corrections for multiple tests (e.g., Bonferroni) but the two-step approach can be conceived of as confirming the significance of correlations that emerged under an exploratory analysis.

negative correlation between tutor reports of student confusion and posttest scores supports this interpretation, as frustration is known to negatively impact learning [11].

Tutor reports of their own frustration, and of student frustration, correlated with student-reported physical demand. That is, the more physically demanding the student felt the task was, the more frustrated the tutor felt and believed the student felt as well. The student rating of physical demand was measured with an item phrased, "How physically demanding was the task?" At first glance, it is unclear whether the students were rating discomfort related to movement/sitting or physiological stress, since (as will be described in Section 6) student report of physical demand did not correlate with measures of posture and gesture. The significant correlation with this report of physical demand may indicate a co-occurring pattern of negative interaction in which the tutor was frustrated and the student was stressed.

It is also worth exploring the cognitive dimensions of the tutors' reporting. The tutors' reports of helping the student complete the task more swiftly and helping the student better master the subject material were both negatively correlated with focused attention, as described above. However, none of the other tutor cognitive reports noted in Figure 4 correlated across the two-phase analysis. Additionally, student performance, as measured by test performance and learning gains, yielded a single correlation between posttest score and tutor report of student confusion. This may indicate that the phenomena evidenced here are related more to implicit perception of cognitive-affective phenomena than to purely cognitive or task-related phenomena. However, the interplay of cognition and affect in task-oriented domains merits further study.

6. MULTIMODAL ANALYSES

The results in the previous section suggest that cognitive-affective states are implicitly communicated in textual dialogue. Examining bodily expressions such as posture and gesture may reveal aspects of affective ground truth related to the implicit affective channel.

Two aspects of posture were used as features in the models reported here. First, variance of the tracked posture points was used as a measure of quantity of motion. Second, the average postural position across a session was used to capture the predominant body position of the student. In addition to these, gesture features include relative frequencies of one-hand-to-face and two-hands-to-face gestures. The set of posture and gesture features is shown in Table 2. The *H*, *M*, and *L* prefixes correspond to the three posture estimation points (Figure 5), while the *All* prefix indicates the sum of all three points. After removing sessions with errorful depth recordings, thirty-one sessions were included in the multimodal analyses.[2]

[2] Four sessions did not have depth recordings due to human error, as recordings were manually initiated. Three sessions were discarded due to extreme postural positions, such as a student leaning far to the side for the majority of the session. Four sessions were discarded due to dark and/or wavy hair that produced persistent noise in the infrared signal, or wearing a baseball cap.

Table 2. Posture and gesture features used in analyses

Feature Set	Feature Names
Averages of posture points	*HAvg*, *MAvg*, *LAvg*, *AllAvg*
Variances of posture points	*HVar*, *MVar*, *LVar*, *AllVar*
Relative frequencies of hand-to-face gestures	*NoGestRFreq, OneHandRFreq, TwoHandsRFreq*

6.1 Correlational Analyses

The first analysis goal was to determine whether the cognitive-affective dimensions investigated earlier also correspond to bodily movements. To accomplish this, correlational analyses were performed between posture/gesture and variables that were involved in significant correlations reported in Section 5. The posture and gesture features in Table 2 were paired with the survey variables in Table 1, and the resulting statistically significant correlations are shown in Table 3 (N=31).

Posture and gesture primarily correlate with survey variables for three cognitive-affective phenomena: student self-report of focused attention, tutor report of student confusion, and tutor report of both student and tutor frustration. Students' report of increased focused attention corresponded to less movement in the lower torso (*LVar*), and to a lower frequency of two-hands-to-face gestures (*TwoHandsRFreq*). These correlations may highlight instances of students leaning forward onto both hands while also moving about the lower torso. Additionally, *LVar* negatively correlated with posttest score, which may illustrate a trend related to lower focused attention.

Table 3. Posture and gesture correlations with survey variables (T=tutor report; S=student report)

First Variable	Second Variable	*r*	*p*
Focused Attention^S	*LVar*	-0.37	0.040
	TwoHandsRFreq	-0.39	0.031
Student Confusion^T (Figure 4, item 15)	*HAvg*	-0.44	0.012
	MAvg	-0.47	<0.01
	LAvg	-0.40	0.026
	AllAvg	-0.48	<0.01
Student Frustration^T (Figure 4, item 18)	*MAvg*	-0.38	0.035
	LVar	-0.37	0.043
	OneHandRFreq	-0.43	0.017
Tutor Frustration^T (Figure 4, item 25)	*HVar*	0.44	0.013
	NoGestRFreq	0.37	0.043
	OneHandRFreq	-0.39	0.029
Posttest Score	*LVar*	-0.40	0.026

Tutor reports of student confusion negatively correlated with average student postural distance. Thus, higher tutor reports of student confusion co-occurred with a more forward student body position. Conversely, farther postural distances co-occurred with lesser tutor reports of student confusion. Similarly, farther mid torso distances co-occurred with tutor reports of student frustration. Taken as a whole, these correlations appear to suggest that forward-leaning postures occur with negative cognitive-affective experience (as perceived by the tutor). Conversely, average postural configurations closer to a straight sitting posture co-occurred with more positive cognitive-affective experience.

Tutor reports of student and tutor frustration both negatively correlated with relative frequency of one-hand-to-face gestures. This is in contrast with the two-hands-to-face gesture, which co-occurred with lower student focused attention. It may be that one-hand-to-face gestures tend to express a positive or thoughtful state, as noted in related literature [13].

6.2 Predictive Models

The correlations presented in Section 6.1 suggest ways in which posture and gesture are associated with student and tutor perceptions of cognitive-affective experience. To further elucidate these relationships, multivariate regression models were built with the significantly correlated variables as predictors. The three survey variables for student focused attention, confusion, and frustration were modeled as outcome variables. Each stepwise linear regression used a conservative 0.05 significance threshold for addition of features.

The regression model for focused attention, shown in Table 4, incorporates two-hands-to-face gestures with variance and average postural position of the lower torso. The model R^2 shows that a moderate amount of the variance in focused attention is explained. Both two-hands-to-face gestures and lower torso variance were negative predictors of focused attention. However, lower torso average distance explains further variance of focused attention as a positive predictor. This model augments the results of the correlational analyses by showing that posture and gesture together combine to predict student focused attention.

Table 4. Stepwise linear regression model for student-reported Focused Attention. Partial R^2 shows the contribution of each feature, while model R^2 shows cumulative model effect.

Focused Attention =	Partial R^2	Model R^2	p
-40.90 * *TwoHandsRF*	0.150	0.150	0.031
-2.90 * *LVar*	0.143	0.293	0.025
0.99 * *LAvg*	0.103	0.396	0.041
23.66 (intercept)	**RMSE** = 10% of variable's range		

The stepwise linear regression model for tutor-reported student confusion, displayed in Table 5, contains a single posture feature that explains a small amount of variance. The absence of additional features shows that the other posture features correlated with student confusion in Table 3 were redundant.

Table 5. Stepwise linear regression model for tutor-reported student confusion.

Confusion =	Partial R^2	Model R^2	p
-0.16 * *AllAvg*	0.231	0.231	<0.01
4.24 (intercept)	**RMSE** = 20.4% of variable's range		

The stepwise linear regression model for tutor-reported student frustration, shown in Table 6, includes relative frequency of one-hand-to-face gestures and lower torso variance as negative predictors. Contrary to the correlational analyses in Table 3, head variance and absence of hand-to-face gestures did not meet the threshold of significance. This model underscores the interplay of posture and gesture, as the addition of lower torso nearly doubles the explained variance.

The regression analyses revealed cumulative effects when posture and gesture were integrated into linear regression models. The model for focused attention incorporated two-hands-to-face gestures and lower torso variance and average distance. Either posture or gesture alone would have explained a small amount of variance, so this demonstrates that the combination of multimodal features such as posture and gesture can improve a predictive model.

Table 6. Stepwise linear regression model for tutor-reported student frustration.

Frustration =	Partial R^2	Model R^2	p
-2.96 * *OneHandRF*	0.182	0.182	0.017
-0.16 * *LVar*	0.155	0.337	0.016
2.86 (intercept)	**RMSE** = 16.2% of variable's range		

The regression model for student confusion did not include gesture, with a small amount of variance explained. However, the regression model built for student frustration included one-hand-to-face gestures and lower torso variance features, resulting in greater explained variance. The root mean squared error of the model for student frustration was less than that of the model for student confusion, as would be expected of the model that explains more variance.

7. CONCLUSION

Although textual communication has limited bandwidth, interactions through the medium still retain high cognitive and affective complexity. If the textual content itself is emotionally sparse, the participants may rely on implicit interpretation of affect. The information relevant to this interpretation may be conceived of as being transmitted through an implicit affective channel. Understanding this implicit affective channel may hold great benefit for systems that aim to interact in naturalistic ways with humans. Toward this end, this paper has presented an empirical study to investigate two aspects of implicit affective communication in textual dialogue: 1) The extent to which the participants converge on shared perceptions of affect, and 2) The ways in which affective ground truth, as captured by depth recordings of gesture and posture, correlates with those affective perceptions even when the bodily movements were not transmitted to the other participant.

This paper has introduced novel posture and gesture recognition algorithms that are robust to occlusions such as desks as a first step toward developing more sophisticated techniques. The automatically recognized posture and gesture features were explored within models that indicate *focused attention*, *physical demand*, and *frustration* were perceived through the hypothesized implicit affective channel accompanying textual dialogue. These results support the notion that an implicit affective channel is at work within computer-mediated textual communication, and that bodily displays of posture and gesture co-occurred with implicit affective communication.

Future investigations of the hypothesized implicit affective channel should seek to identify specific processes through which individuals are able to interpret affect in computer-mediated textual dialogue. For instance, analyses based on fine-grained temporal features of computer-mediated textual dialogue may reveal relationships between perceived affect and how participants construct messages. Additionally, bodily expressions that correlate with aggregate cognitive-affective measures may be explored within discrete time windows to provide a more dynamic view of affective phenomena that occur during textual dialogue. Machine learning techniques may also be applied to identify complex patterns of interaction that may shed light on underlying

human processes. Such investigations may lead to future textual dialogue systems that leverage the implicit affective channel.

ACKNOWLEDGEMENTS

This work is supported in part by the North Carolina State University Department of Computer Science along with the National Science Foundation through Grant 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

[1] Baltrusaitis, T. et al. 2011. Real-Time Inference of Mental States from Facial Expressions and Upper Body Gestures. *Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition*, 909-914.

[2] 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.

[3] Csikszentmihalyi, M. 1990. *Flow: The Psychology of Optimal Experience*. Harper-Row.

[4] Derks, D. et al. 2008. The Role of Emotion in Computer-Mediated Communication: A Review. *Computers in Human Behavior*. 24, 3, 766-785.

[5] 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*, 31-40.

[6] Forbes-Riley, K. and Litman, D. 2011. Benefits and Challenges of Real-Time Uncertainty Detection and Adaptation in a Spoken Dialogue Computer Tutor. *Speech Communication*. 53, 9-10, 1115-1136.

[7] Graesser, A.C. and Olde, B.A. 2003. How Does One Know Whether a Person Understands a Device? The Quality of the Questions the Person Asks When the Device Breaks Down. *Journal of Educational Psychology*. 95, 3, 524-536.

[8] Hart, S.G. and Staveland, L.E. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. *Human Mental Workload*. P.A. Hancock and N. Meshkati, eds. Elsevier Science. 139-183.

[9] Kapoor, A. et al. 2007. Automatic Prediction of Frustration. *International Journal of Human-Computer Studies*. 65, 8, 724-736.

[10] Kiesler, S. et al. 1984. Social Psychological Aspects of Computer-Mediated Communication. *American Psychologist*. 39, 10, 1123-1134.

[11] Kort, B. et al. 2001. An Affective Model of Interplay between Emotions and Learning: Reengineering Educational Pedagogy-Building a Learning Companion. *Proceedings of the IEEE International Conference on Advanced Learning Technologies*, 43-46.

[12] Mahmoud, M. et al. 2011. 3D Corpus of Spontaneous Complex Mental States. *Proceedings of the International Conference on Affective Computing and Intelligent Interaction*, 205-214.

[13] Mahmoud, M. and Robinson, P. 2011. Interpreting Hand-Over-Face Gestures. *Proceedings of the International Conference on Affective Computing and Intelligent Interaction*, 248-255.

[14] Marcoccia, M. and Atifi, H. 2008. Text-Centered Versus Multimodal Analysis of Instant Messaging Conversation. *Language@Internet*. 5.

[15] McDuff, D. et al. 2011. Crowdsourced Data Collection of Facial Responses. *Proceedings of the 13th International Conference on Multimodal Interfaces*, 11-18.

[16] McNeill, D. 2005. *Gesture & Thought*. The University of Chicago Press.

[17] Neviarouskaya, A. et al. 2011. Affect Analysis Model: Novel Rule-Based Approach to Affect Sensing from Text. *Natural Language Engineering*. 17, 01, 95-135.

[18] Neviarouskaya, A. et al. 2007. Analysis of Affect Expressed through the Evolving Language of Online Communication. *Proceedings of the 12th International Conference on Intelligent User Interfaces*, 278-281.

[19] Neviarouskaya, A. et al. 2007. Textual Affect Sensing for Sociable and Expressive. *Proceedings of the International Conference on Affective Computing and Intelligent Interaction*, 220-231.

[20] 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.

[21] Pantic, M. et al. 2006. Human Computing and Machine Understanding of Human Behavior: A Survey. *Proceedings of the 8th International Conference on Multimodal Interaction*, 239-248.

[22] Perry, M.S. and Werner-Wilson, R.J. 2011. Couples and Computer-Mediated Communication: A Closer Look at the Affordances and Use of the Channel. *Family and Consumer Sciences Research Journal*. 40, 2, 120-134.

[23] Quek, F. et al. 2002. Multimodal Human Discourse : Gesture and Speech. *ACM Transactions on Computer-Human Interaction*. 9, 3, 171-193.

[24] Sanghvi, J. et al. 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*, 305-311.

[25] Schuller, B. et al. 2011. Recognising Realistic Emotions and Affect in Speech: State of the Art and Lessons Learnt from the First Challenge. *Speech Communication*. 53, 9-10, 1062-1087.

[26] Walther, J.B. 1992. Interpersonal Effects in Computer-Mediated Interaction: A Relational Perspective. *Communication Research*. 19, 1, 52-90.

[27] Walther, J.B. 2007. Selective Self-Presentation in Computer-Mediated Communication: Hyperpersonal Dimensions of Technology, Language, and Cognition. *Computers in Human Behavior*. 23, 5, 2538-2557.

[28] Woolf, B.P. et al. 2009. Affect-Aware Tutors: Recognising and Responding to Student Affect. *International Journal of Learning Technology*. 4, 3-4, 129-164.

[29] Zeng, Z. et al. 2009. A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 31, 1, 39-58.