5-10-2013

Measuring attention using Microsoft Kinect

Darren Stanley

Measuring Attention using Microsoft Kinect: Master's Thesis

Darren Stanley
Department of Computer Science
College of Computing and Information Sciences
Rochester Institute of Technology

May 10, 2013

Abstract

The transfer of knowledge between individuals has increasingly become achieved with the aid of interfaces or computerized training applications. However, computer based training currently lacks the ability to monitor human behavioral changes and respond to them accordingly.

This study examines the ability to predict user attention using features of body posture and head pose. Predictive abilities are assessed by an analysis of the relationship between the measured posture features and common objective measures of attention, such as reaction time and reaction time variance. Subjects were asked to participate in a series of sustained attention tasks while aspects of body movement and positioning were recorded using a Microsoft Kinect. Results showed support for identifiable patterns of behavior associated with attention while also suggesting the complex inter-relationship of measured features and susceptibility of these features to environmental conditions.

Signatures

I, Darren Stanley, do hereby submit this thesis in partial fulfillment of the requirements for the degree of Master of Science in Computer Science. It is approved by the committee memebers below.

Darren Stanley

Dr. Joe Geigel
Committee Chair

Dr. Reynold Bailey
Reader

Sean Strout
Observer

Contents

Chapter 1

Introduction

In today's world, computers are an integral part of our everyday life. This is especially true in the educational domain, where computers can perform complex calculations and quickly access resources from around the world. With the help of technology, teachers are able to present material in real time to students across the world. Moreover, computer training programs are able to be run over and over again in any place and at any time. We have done a remarkable job at creating utilities that aid students, and any other individual seeking knowledge, to obtain it. Whether participating in a webinar, watching a pre-recorded training video, or even participating in a virtual distance learning classroom [8], most of us have benefited from the opportunities computer assisted training provides. However, the increased interaction with computers does introduce some problems. Students who are less interested in the topic or less self-motivated can find serious disadvantages to computerized training. Prebuilt applications, while widely accessible, may not be able to pace the material appropriately for the student, possibly frustrating students who need the material presented slower and those who would like to move at a faster rate. In distance learning environments, it can be difficult for the teacher to be effective since there can be little or no feedback from the audience. The introduction of a computer interface limits the adaptability of the training by removing or ignoring the human factors that serve as indicators of the learner's interest and involvement. For this reason automated tutoring applications are a long way from achieving the results of direct human tutoring. The next step to improving the use of technology in education is the ability to create computer assisted training that is able to recognize and react to human behavioral changes.

The idea of monitoring human behavior computationally is not new. Over the years, different techniques have been applied to affective and cognitive state detection. Two of the most well known approaches are facial recognition algorithms and brain-wave recordings through electroencephalography(EEG). Significant progress has been made in both fields, but they each suffer from problems that have prevented them from wide-spread adoption.

The first limitation is cost. Both facial recognition and EEG technologies require expensive sensors that are not affordable to the general public. This limits both the exposure to the technology and its applications. The second problem is that of invasiveness. Sensors that need to be worn on the body or limit the mobility of the learner are more difficult to introduce and gain acceptance.

Recognizing these deficiencies, researchers have begun to look for alternative approaches. Until recently, the analysis of posture and gross body movement to determine cognitive states has been limited. This is largely due to the fact that posture was difficult to measure without using invasive or expensive technology. Advances in technology, such as the release of the Microsoft Kinect, have provided lower cost, non-invasive tools for capturing posture data. [5] looked at using pressure sensors on seats and [19] used video to capture learners seated posture as non-invasive measures of affect and user engagement. These techniques showed that useful information can be obtained from body posture alone and without hindering the learners mobility. [18] took the lessons learned from [19] and [5] and applied them to data collected from the Microsoft Kinect.

This research explored the use of postural data points, collected by Microsoft Kinect, for detecting user attention. Using data collected form a Microsoft Kinect, this study monitored the posture and movement of an individual during a computerized attention assessment exam. This work compliments the previous work done by [5] and [19] towards recognition of engagement and affect. In addition, this research explored the ability to use Kinect skeletal tracking data for posture recognition. Specifically, this work looked to analyze the Kinect's ability to provide useful skeletal frame data for posture recognition, determine the usefulness of posture data for detecting attentive states, determine which features, if any, collected from the Kinect are useful in measuring attention, and build a framework to correlate skeletal wireframe data with existing measures of attention.

Chapter 2

Background

2.1 Attention

> Every one knows what attention is. It is the taking possession by the mind, in clear and vivid form, of one out of what seem several simultaneously possible objects or trains of thought. Focalization, concentration, of consciousness are of its essence. It implies withdrawal from some things in order to deal effectively with others, and is a condition which has a real opposite in the confused, dazed, scatterbrained state which in French is called distraction. [13]

Despite William James' famous observation about attention, the definition of attention is often context specific. In general, attention is a type of focus that allows us to perceive certain objects or thoughts above others. The object that we are focusing on is commonly called the stimulus and we usually refer to the act of choosing on which stimulus to focus selecting.

We can delineate attention into two broad categories defined by the selection mechanism used. *Overt attention* involves an attention mechanism that selects specific kinds of neural processing by physically moving the sensory organs [9]. Overt attention is generally associated with external stimuli and is by far the most commonly studied form of attention because the movement of the sensory organ provides a measurable value for analysis. The other category of attention is *covert attention*, which is concerned with attention mechanisms that do not involve explicit movement of the sensory organs [9]. Covert attention can orginate from an internal stimulus or

be a further contemplation on a previously collected external stimulus.

Overt attention is usually described in terms of what sensory organ perceives the stimulus, for example visual attention or auditory attention. In general when attention is being described by a sensation (visual, auditory, etc) there is an implication that the source of the stimulus is external.

The *orienting response* is the changing of orientation for a particular sensory organ in response to the recognition of an external stimulus.

Since covert attention refers to the idea that attention can be directed inward or reflected back on one's own thoughts, it is not typically identifiable by the orientation of an individual towards a stimulus but can sometimes be identified by the abscence of an expected orientation. The lack of an orienting response to a given stimuli can indicate either a malfunction of the sensory organ or a strong attention to some other stimuli. If there is no orienting response for any of the sensory organs, it may imply the subject is directing his or her attention inwardly, or covertly attending.

Most research aimed at understanding or measuring attention is geared toward overt attention. In particular, most experimentation involves either visual or auditory stimuli. One exception to this is the EEG, which focuses on signals sent to the brain and is capable of detecting focus directed externally or internally. However, even when using the EEG to measure attention, it is generally desireable to control the stimulus and therefore a stimulus external to the subject is typical.

2.1.1 Measuring Attetion

A great deal of research has been done on how to effectively measure attention. Most of this work involved studying individuals with learning disabilities. The ability to focus on and attend to a given task are critical aspects of learning. In order to study the different demensions of attention, a number of tests have been created to specifically target indiviual situations. Often a comprehensive analysis of a subject's attention requires performing multiple of these tests.

Alertness

An alertness task is any test where a subject is asked to respond to a particular stimulus when it is presented. The objective of such a task is to measure one's ability

to remain focused and alert over time. Over time it is expected that one's level of alertness will decrease. The rate at which that measure decreases, and the consistency of the decline can be used to describe how well a person can sustain attention. A high variance in reaction time or large number of omissions might indicate that one has trouble maintaining the same level of alertness or that their attention drops more quickly over time than average. The measure of reaction time alone may be misleading, since slow reaction time alone does not mean that the subject was not paying attention. It is possible that the stimulus is too complex to quickly identify or that the subject is struggling to identify the stimulus for some other reason. Comparing the reaction time of an individual against the group provides insight into whether the subject is able to recognize the stimulus as effectively as the rest of the group. Looking at the variance (amount of change) in the reaction times for an individual allows us to determine how quickly the subject's attention dissipated or how often their attention strayed. A linear decrease in reaction times expresses waning attention while unpredictable variance in reaction time expresses a tendency towards attentional shifts. The number of omission errors is similar to the reaction time variance; it is expected that over time an individual will experience a higher frequency of missed signals, but a steep increase or frequently sporadic misses might indicate trouble maintaining attention.

Often when collecting measurements on alertness, both tonic and phasic alertness are measured. Tonic alertness is a measure of alertness when the subject is asked to respond to a stimulus without prior warning. The subject is presented with a stimulus randomly over a period of time and asked to respond each time the stimulus is present. When measuring phasic alertness, the subject is still asked to respond whenever a stimulus is presented over a period of time; however, during the test the subject is always presented with a warning stimulus before the target stimulus is presented. The warning stimulus gives the subject a chance to refocus their attention before the target stimulus is presented.

An alertness task is good at measuring how well an individual can sustain attention in terms of how consistent that attention is sustained and how long it is sustained.

Vigilance

A vigilance task is a type of alertness task where the target stimulus is presented infrequently over a relatively long and continuous period of time. The vigilance task is useful for determining if an individual is able to remain focused on a task of particu-

larly low interest. An individual who performs well on a standard alertness task may not necessarily do well with vigilance tasks since the vigilence task requires greater patience and self control. In addition to the metrics used to measure alertness, vigilance also measures the number of commission errors (an error in which the subject erroneously responds to a non-target stimulus). The number of errors of commission are likely to be higher for subjects who exhibit more impulsive behavior. Vigilance tasks are representative of many real life tasks where the stimulus is unknown and occurs at an unknown time. For example, a student listening to a lecture does not know when the most useful information will be presented or even what the most useful information is going to be. For many tasks where the target stimulus is unknown (such as in the lecture example) one must be vigilant and attend to all stimulus and be able to decide what is important.

Divided Attention

Divided attention tasks are another type of task that is often used to asses a subject's likelihood of having ADHD. In a divided attention task, a person is asked to respond to multiple target stimuli simultaneously. Children suffering from ADHD have a tendency to perform poorly on divided attention tasks and are more likely to commit the entirety of their attention towards a single task. The metrics used for measuring vigilance – reaction time, variance in reaction time, number of omission, and number of commission errors – are also used for divided attention tasks.

Visual Scanning

Visual scanning involves an image or series of images that may or may not contain a predefined target stimulus. The subject is presented with these images and asked to indicate if the stimulus is present. Visual scanning is a selective attention task that measures inhibition or impulsivity. Participant are asked to respond to either the presence or absence of a designated stimulus. [21] Individuals with higher than normal inhibition will have a longer reaction time and more errors of omission than average test subjects. Individuals with impulsive tendencies will have a shorter than expected reaction time and a higher than normal number of commission errors. This type of task is very useful for measuring impulsivity, but not as useful in assessing inattention.

Incompatibility

The incompatibility test is used to measure how well an individual filters irrelevant data. Participants are asked to respond differently to multiple target stimuli where some of the stimuli contain contradictory content. For example, the subject may be presented with the word GREEN written in red text. How the subject responds to these contradictory cases measures how well the individual filters irrelevant data. This task also uses reaction time, reaction time variance, and number of commission errors to analyze the subject. A subject who is inattentive may be expected to miss the same ratio of target stimuli for both compatible and incompatible presentations while a subject who struggles to select only the relevant information will have a higher number of errors on incompatible presentations. This is valuable for determining if the subject is suffering from inattention or some other learning disability.

Flexibility

The flexibility task measures the flexibility (ability to change) of focused attention by a mental alteration between two sets of targets. [22] This test helps to expose subjects who have difficulty context switching, but does not necessarily indicate if the subject has trouble maintaining attention. In one example of a flexibility task, a subject is placed in front of a computer screen with each hand placed on a different button. The participant is instructed to respond by alternately pressing the button that was on the same side of the screen as the letter, and then pressing the button that was on the same side of the screen as the number. After each response, a new letter and number appears, randomly assigned to either side of the screen. [21]

Cross-Modal Integration

Cross-modal integration tasks measure one's ability to incorporate information from multiple stimuli (in different modalities) in real time. This is useful for distinguishing the difference between a subject who is inattentive from a subject who has difficulty integrating information from multiple sources. Usually, the subject is presented with two stimuli simultaneously, each belonging to a different modality (vision, hearing, etc), and asked to respond when a certain criteria (involving aspects of both stimuli) is met. The response time, response time variance, number of omission errors, and number of commission errors are recorded as part of this task. A suggested cross-modal integration task involves presentation of either an up or down arrow on a

display while simultaneously playing either a high or low frequency audio tone. The subject is asked to press a response button when the up arrow is displayed along with the high frequency sound or when the down arrow is displayed along with the low frequency sound. [22] [21]

2.1.2 Continous Performance Task

The Continuous Performance Task (CPT) is a common method for clinically assessing ADD [17]. The CPT is a form of sustained attention task that places the subject in front of a computer monitor and asks them to respond to a particular stimulus when it is presented on the screen. Typically the test will run for 14 minutes in order to measure the subjects performance over time and allow for subjects to become impulsive or inattentive.

Most CPTs measure response time, errors of omission, and errors of commission. Omission errors occur when the subject does not respond to the target stimulus and indicate the individual is not paying attention, while commission errors occur when the subject incorrectly responds to a non-target stimulus which indicate the individual is reacting impulsively. Individuals with inattentive and impulsive tendencies will record a higher number of these errors and suffer from inconsistent response times.

2.2 Microsoft Kinect

The Microsoft Kinect was released in November 2010 and was originally designed as an accessory for the Xbox 360 gaming system. The main function of the Kinect is to provide gross body movement capture in a non envasive and affordable manner. This allowed the development of games that required no controller and could be played entirely by user gestures and body movement.

The computer science community quickly recognized the potential of the Kinect for enhancing human computer interaction (HCI) in numerous fields. An open source project was started to provide an API (OpenNI) for working with the Kinect from a computer rather than the Xbox console. Based on the success of this API and the high demand for an official API, Microsoft released an API in June 2011. The official API provides access to depth, RGB, and articulated figure data. A new version of

the Kinect, specifically designed for use with a computer, was released in February 2012 along with updated hardware and an enhanced API (Kinect 1.5 SDK).

2.2.1 Technical Specifications

Kinect for Windows can run on any machine running Windows 7 or greater with a modern processor and at least 2GB of RAM [4].

The Kinect sensor (also called a Kinect) is a physical device that contains two cameras, a microphone array, and an accelerometer and comes with a software pipeline that processes color, depth, and skeleton data.

Figure 2.1: Image of Kinect sensor.

The Kinect sensor can identify human figures within a 58 degree horizontal and 44 degree vertical field of view. The Kinect can operate in either of two modes 'default' or 'near'. The default mode has been supported since the original SDK release and is used for tracking standing figures that are within 1.2 and 3.5 meters away. Twenty joint positions are tracked on up to 2 individuals simultaneously. While the Kinect can only track skeletal data for 2 figures it is capable of recognizing up to 6 within its range of view. Near mode was added in SDK v1.5 and is designed for seated figures 0.8 to 2.5 meters from the sensor. Near mode only tracks 10 joint positions ignoring those of the lower body which are generally hidden beneath a desk or other obstacle.

This study only minimally takes advantage of the audio capabilities of the Kinect by tracking the existence of noise above a threshold coming from the direction of the subject. This is made possible by the Kinects microphone which can be directed at a target in 10 degree increments up to 50 degrees in either direction, or 100 degrees range total. The sensor and allows 20dB of ambient noise cancellation [4] which will help rule out false positives and correctly identify when the subject is speaking.

As part of the Kinect SDK, Microsoft released a face tracking API which utilizes the depth and RGB camera to track 87 distinct facial points. This allows developers to calculate facial expressions or drive the movement of a 3D avatars mouth and eyes. The face tracking API also tracks the pitch, roll, and yaw of a persons head within 90 degrees from center.

Chapter 3

Related Work

The concept of using posture to measure cognitive and emotional states has appeared in various forms and different fields of study throughout the years. Existing work has been done to measure attention using aspects of posture and observable behavior and a good deal of research has recently been directed toward automated techniques for detecting user affect and engagement using postural data.

3.1 Objective Measures of Attention Using Posture

3.1.1 Behavioral Coding Systems

One method that has been used to measure attention and impulsiveness in children is to use a behavioral coding system. Coding systems provide a well defined set of observable behaviors that an observer can use to 'score' a child's attention level. The use of predefined behavorial definitions helps to eliminate bias and helps ensure that all observers are using the same criteria to asses an individual. Studies have shown that behavioral coding systems, such as the Abikoff coding system, which provides a set of observed behaviors to score the likelihood a child may be experiencing a learning disability, are effective tools for objectively identifying children with various learning disorders [12].

Behavioral coding systems are closely related to this study because both rely on the

idea that common, observable behaviors which are objective can be used to gauge levels of attention. The success of coding systems like Abikoff provide strong support that it is possible to build an automated system for detecting attention. In order to be successful the system needs to be able to correctly identify the defined set of objective behaviors and then build a score from its observations.

3.1.2 Gaze Detection

As discussed earlier, overt attention requires the use of a sensory organ to process information about the stimulus. Aspects of our attention that are directed at visual targets can be measured by external observation of the sensory organ used. That is, an observer can watch where our sensory organ (in this case the eyes) is directed and infer that the subject is attending to the stimulus. A common scenario where we might want to measure attention is a student during a classroom lecture. During this scenario we can say there are two target stimuli, the teacher, who is speaking, and the whiteboard, which contains supporting content for the lecture. We can easily identify if the student is visually attending to either target by analyzing their gaze direction and determining if they are looking at either target. Determining if the student is listening is more complex, but those who are not visually attending to a presentation are less likely to be paying attention to it auditorially. This idea has been used as an approximation of a student's attentiveness during automated engagement detection. [2] uses video recording with face recognition to calculate the direction students are looking in order to classify them as attending or not-attending. This feeds into a larger behavior analysis framework which calculates the overall engagement level for each student.

3.2 Predicting Affect and Engagement Through Posture

Promising work has been done identifying user engagement based on posture analysis. Basic affective states and engagement levels were successfully identified by a system measuring posture with a collection of pressure sensors placed on both the seats and backs of chairs [5]. Using this technique [5] was able to show that measuring posture and gross body movement could be used for predicting the engagement level of a user. The total quantity of the individuals movement along with the location

and amount of pressure placed on the different regions of the seat were also found to be revealing signs of affective state. In other research, [19] used video analysis to study the posture of children playing a game with a robotic interface and found that the angle and curvature of the child's back could be used to identify the level of engagement that child experienced during the gaming session. Early exploration with the Kinect has begun to show it could be an effective tool for collecting posture related data. [18] used depth information reported from the Kinect to calculate a vector consisting of the lower torso, mid torso, and head distances from the sensor. This vector was used as an approximation of the subjects body lean and was shown to support the findings described by [19].

3.2.1 Pressure Sensors

Detection of learners affect from gross body language collected from pressure sensors located along the seat and back of a chair was conducted in comparison with other communicative techniques, such as facial recognition and dialogue analysis [5]. This research used the pressure sensors to gather feature vectors for use with traditional classifier algorithms, such as Bayesian classifiers, functions, instance-based learners, meta classifiers, rules, and trees. Specifically, the average amount of pressure placed on certain regions of the sensor along with the magnitude and direction of pressure changes was used feed these classifiers.

Participants interacted with an automated tutoring application called AutoTutor while being observed by a panel of judges. The judges were responsible for classifying the subjects affect throughout the training session. After each session the judges ratings were combined with the collected feature vectors to be used as a training set for the classifiers. Results showed that with body pressure features alone, conventional classifiers were moderately successful at discriminating the affective states of boredom, confusion, delight, flow, and frustration. Boredom and flow detection was significantly higher than confusion and delight, which appear to be more facially expressive. [5] claim their results indicate the face is not the most significant communicative channel for some of the learning centered affective states, such as boredom and flow.

3.2.2 Body Lean

Some interesting work has been done connecting the angle of lean towards a stimulus to the level of engagement the subject is experiencing. Findings show that engaged users tend to increase their proximity to a target with increasing interest. This is generally achieved by leaning towards the target. Several research efforts have explored techniques for capturing this information.

[5] indirectly captured this information from pressure sensors placed on a chair. When the subject is leaning back in the chair more pressure is recorded by the sensors and when the subject leans forward (toward the target stimulus) less pressure is recorded.

Previous work, specifically that of [5], has identified that when subjects are bored they have a tendency to lean back, while during delight and flow (affective states correlated with engagement) students tend to lean forward. Students also lean forward when experiencing confusion and frustration, but at a lesser inclination All of these states suggest a high probability of attention focused on the target. It seems likely that students would focus more attention toward the target in both scenarios. During frustration and confusion the student focuses more as an attempt to resolve the source of frustration while during delight and flow the student attends because of their interest in the topic. [5] were able to identify when a subject was leaning forward using a series of pressure sensors (discussed earlier) placed on chairs. If the subject was leaning back in the chair pressure values increase on the sensors located on the back of the chair and when the subject leans forward pressure values decrease. This provides a good measurement of when a student is leaning forward.

A more detailed discussion of body lean can be found in the work of [19] who demonstrated that both the angle the subject is leaning and the curvature of the subject's back are important indications of user engagement. In their work, engagement is defined as the value that a participant in an interaction attributes to the goal of being together with the other participant(s) and continuing the interaction []. In order to measure that angle of lean and curvature of the back, a more advanced technique was required. Computer vision techniques were applied to extract postural features from videos of the subject interacting with a robotic game opponent. In order to determine which measurements were important for detection of engagement, a panel of judges were used to rate the subjects engagement level during the recorded session. Those ratings were then correlated with postural features extracted from the video during the same time interval.

3.2.3 Measuring Engagement Using Kinect

There is also existing research which has identified the possibilities opened up by the release of the Microsoft Kinect for measuring engagement levels. So far, this work has not been able to take advantage of the articulated data provided by the Kinect, but has shown even the raw data can be used to effectively gauge engagement.

3.2.4 Discussion

All of the measures of affect and engagement discussed so far rely on subjective analysis. The two techniques discussed in this review for capturing affect and engagement are self rating and analysis by a trained judge. In the first scenario the subject is given a questionnaire about their experience and asked to recall their emotions and state of mind during the examination. The second technique requires the ability to obtain subject matter experts and have them watch the subject during the experiment and record their perception of the individuals state of mind. In contrast to the measurement of engagement, the study of attention has made greater use of more reproducible and unbiased measurements and assessment techniques.

There are, however, a number of similarities between measuring engagement and attention. Psychologists have long acknowledged the relationship between observable physical behavior and attention. In fact, this is the foundation of *overt* attention mechanisms. Another example is the identification of children with ADD. Often children are first suspected of having ADD by their parents or teachers based on observed behavior patterns. Indeed, some of the earliest techniques for identifying ADD were behavioral checklists filled out by an expert or someone close to the child. This is analogous to the self assessment exams given to subjects after a training session.

Recently, measuring user engagement and affect has received increased attention because of the desire to improve human-computer user experience and maximize the potential of computer augmented training technologies. There are a number of studies showing the benefit of engaging the user given the relationship between affect, engagement and attention, even if these relationships are not yet well defined. Affect is likely to effect the willingness of a subject to attend to a particular stimulus while engagement is partially defined by the subjects attendance to a stimuli. What needs to be done is to look at the techniques used for measuring attention combined with the methods used recently to non-invasively measure affect and engagement.

3.3 Hypothesis

The purpose of this study is to determine if a system can be built to recognize both inattentive and attentive behaviors in order to use that information as feedback for dynamic improvement. Some behaviors are obvious indications of inattention while others are more subtle. For example, when a person is not looking at a computer screen during a presentation, it is obvious they are not visually attending to the target. Using the measures described in the methodology section, this study seeks to categorize inattentive and attentive behaviors.

More specifically, this study seeks to determine if common behavioral patterns and postures can indicate the attentiveness of an individual. To accomplish this, Kinect hardware, which is capable of recording articulated data over time, was employed to identify such movements and postures. The Kinect has the additional benefit of being cost effective and non-invasive.

The study's hypothesis was that measures reported as having a high correlation to engagement in [10] will also positively correspond to when the subject is actively attending during the exam. The body lean and side lean measurements are an attempt to use skeletal data directly to replicate the results of the head and torso distance measures in previous research. It was also expected that head gaze will positively correlate with subjects' attention and the more movement recorded, the more likely in-attention will be witnessed.

Behavior	Type	Measures
Not Looking	In-Attentive	Gaze Direction
Fidgety/Restless	In-Attentive	Gross Body Movement
Fatigue/Tiredness	In-Attentive	Slouch / Posture
Engagement	Attentive	Body Lean, Head Proximity
At Attention	Attentive	Sitting up-right posture

Table 3.1: Summary of suspected (in)attentive behaviors

Chapter 4

Methodology

4.1 Overview

This study attempted to measure the level of attention for test subjects participating in a sustained attention task. Each subject participated in well known attention tasks in order to provide an objective measure of the subject's attentiveness. This measure was then compared with different measured features of body posture obtained from a Microsoft Kinect. Each test subject participated in 3 Continuous Performance Tasks (CPTs) for a total time of 1 hour. The tests were 14 minutes, 24 minutes, and 17 minutes long and run back to back with a one minute break between.

4.2 Subjects

The study was conducted on 20 normal adult volunteers between the age of 23 and 45, with no known attention defficiences. No requirements were placed on sex or age for this experiment, but the majority of test subjects were male and in their late 20s or early 30s.

4.3 Measures

Measures for this experiment were the results of three sustained attention tasks administered to the test subjects along with corresponding postureal data, collected by a Kinect.

Two of the tasks chosen were Continuous Performance Tasks (CPT), since CPTs tend to be longer than other attention tasks and have been shown to be a good measure of sustained attention. Since the test subjects were normal adults, they were expected to have well developed attention spans compared to young children and therefore less likely to commit errors. Providing a longer, more tedious exam than typically administered, offered a greater opportunity to capture lapses in attention.

Although many versions of the CPT have been developed, the basic methodology of these tasks remains consistent with that of the original. Subjects are presented with a variety of stimuli, which are displayed on a screen for a short duration and are instructed to respond to a predefined "target" stimulus. A number of values are often recorded in these tasks, including omission errors and commission errors. In addition, response times for correct detections and for various commission errors are recorded in an effort to better measure problems of inattention and impulsivity [11].

4.3.1 PEBL

A test battery was created from three sustained attention tasks provided by the Psychology Experiment Building Language (PEBL) in order to measure different features of attention during the experiment. PEBL is an open source language designed for building psychology tests and comes equipped with implementations for a number of common psychological tests including several CPTs. The open source nature of PEBL makes it easy to modify tests to suit individual measurment requirements. For this experiment the PEBL Continuous Performance Test (PCPT), Test of Attentional Vigilance (TOAV), and the PEBL Perceptual Vigilance Task (PPVT) were used.

PCPT: PEBL Continuous Performance Test

Description: This test is an implementation of the well known Conner's Continuous Performance Task (CCPT). This is a (not-X) CPT, which means the subject is

asked to respond every time a letter is presented that is not the letter X. Stimuli are chosen randomly with an 'X' appearing 10 percent of the time. During the PCPT the subject is presented with 360 letters at different inter-stimulus intervals (ISI). The entire test takes approximately 14 minutes. Each stimulus is presented for 250 milliseconds. The test is divided into 24 blocks with each block consisting of 20 trials. The ISI value changes at the beginning of each block, cycling between 1, 2, and 4-second delays. Six ISI cycles are completed during the test. PCPT measures 1) omissions 2) commissions 3) Hit RT 4) RT Std Error 5) Hit RT Block Change and 6) Hit SE Block Change.

Output: PCPT outputs two files: 1) A CSV data file containing features for each trail and 2) a human readable text file report summarizing the results. For a complete description of the PCPT output fields and example documents “see appendix B on page 68”.

Filename Pattern: pcpt-[id].csv
Delimiter: comma (,)

Field #	Column Header	Description
1	sub	Subject Identifier.
2	block	Block Number.
3	trail	Trial Number.
4	cond	Delay (ms) before the stimulus was presented.
5	targ	Target Presented. Target=[A-U], Non-Target=[X].
6	responded	Response to Target Present. 0=No, 1=Yes
7	corr	Correct Response. 0=No, 1=Yes
8	time	Time of Target Presentation (ms from start).
9	rt	Reaction time in milliseconds. -1=No Response

Table 4.1: PCPT raw data file format.

TOAV: Test of Attentional Vigilance

Description: This is the PEBL implementation of the visual version of the well known Test of Variables of Attention (TOVA) continuous performance task. This test runs for 24 minutes, during which the subject is presented with two stimuli occurring at random intervals. Of the two stimuli, one is designated as a target and one is designated as a non-target. The subject is asked to respond to the target

```
--------------------------------------------------------
Statistic        ISI:   1000 2000 4000 Pooled
--------------------------------------------------------
Correct Trials    114/120 111/120 110/120 335/360
Correct Targets   112/112 107/108 108/108 327/328
Correct Foils     2/8 4/12 2/12 8/32
Target Acc Rate   1 0.991 1 0.997
Foil Acc Rate     0.25 0.333 0.167 0.25
Commission Errors 6 8 10 24
Ommission Errors  0 1 0 1
Correct RT Mean   329.37 363.75 419.73 370.46
Correct RT SD     71.35 89.33 75.62 87.4
Error RT Mean     317.33 337.63 426.5 369.58
Error RT SD       41.49 42.19 67.4 72.73
Sensitivity (d')  0 -1.924 0 -2.068
Bias (beta)       0.804 0.764 0.858 0.806
--------------------------------------------------------
```

Table 4.2: Example PCPT report statistics.

whenever it is presented on the screen by pressing a defined key on the keyboard. This test includes two sections. During the first section, known as the infrequent condition, targets randomly occur once for every 3.5 non-targets. This pattern is reversed for the second half of the task, with targets appearing 3.5 times for every one non-target. This task measures Response Time, Response Time Variability, Performance Deterioration Rate, Errors of Omission, Errors of Commission, Post Commission Response Time, Multiple Responses, and Anticipatory Responses.

Output: TOAV outputs two files: 1) A CSV data file containing features for each trail and 2) A human readable text file report summarizing the results. For a complete description of the TOAV output fields and example documents “see appendix D on page 74”.

PPVT: PEBL Perceptual Vigilance Task

Description: This is an implementation of a Psychomotor Vigilance Task (PVT). It runs for 10 minutes and randomly presents a target stimulus to the subject at

Filename Pattern: toav-[id].txt
Delimiter: whitespace (tab, space)

Field #	Column Header	Description
1	sub	Subject Identifier.
2	trail	Trial Number.
3	targ	Target Presented. 0=Non-Target, 1=Target.
4	toofast	Response too fast. 0=No, 1=Yes
5	responded	Response to Target Present. 0=No, 1=Yes
6	corr	Correct Response. 0=No, 1=Yes
7	mult	Multiple Responses to Target. 0=No, 1=Yes.
8	time	Time of Target Presentation (ms from start).
9	rt	Reaction time in milliseconds. -1=No Response

Table 4.3: TOAV raw data file format.

```
---------------------------------------------------
Statistic            Half 1     Half 2     Pooled
---------------------------------------------------
Total Trials        320         320         640
Correct Targets     58          222         280
Correct Foils       239         54          293
Correct Trials      297         276         573
Commission Errors   9           18          27
Ommission Errors    14          26          40
Correct RT Mean     560         424         452
Error RT Mean       593         378         449
RT Mean             564         421         452
RT SD               189         131         157
Anticipations       0           1           1
Multiple Responses  1           2           3
---------------------------------------------------
```

Table 4.4: Example TOAV report statistics.

long intervals. This test is used to identify and record lapses (which are defined as reaction times slower than 500ms).

The test battery collects measurements for a) omissions b) commissions c) Hit RT d) RT Std Error e) Hit RT Block Change f) Hit SE Block Change and g) Number of Lapses.

Output: PPVT outputs two files: 1) A CSV data file containing features for each trail and 2) A human readable text file report summarizing the results. For a complete description of the PPVT output fields and example documents "see appendix C on page 71".

Filename Pattern: ppvt-[id].txt		
Delimiter: single space		
Field #	**Column Header**	**Description**
1	sub	Subject Identifier.
2	block	Block Number.
3	trail	Trial Number.
4	ISI	Delay before target presentation.
5	ISIbin	ISI 'bin' by 1000 ms intervals. [1000-9000]
6	abstime	Time of Target Presentation (ms from start).
7	rt	Reaction time in milliseconds. -1=No Response
8	type	Categorization of the reaction time. 1=Too Fast, 2=Typical, 3=Lapse, 4=Sleep Attack

Table 4.5: PPVT raw data file format.

Because normal, non-inattentive adults tend to have well developed attentional skill, all CPTs suffer from a lack of omission errors when working with older individuals [11]. Nevertheless, the CPT is one of the best measures for attentional features in normal adults. As such, this study focused on using the CPTs' measures for reaction time and reaction time variance for identifying inattention in normal adults [1].

Also, due to the concern of only collecting data from attentive individuals, for the PCPT test; subjects were divided into two groups. The first group participated in the PCPT test as typically administered, without any modifications.

The second group was given a modified version of the PCPT test which required them to attend to a secondary computer screen and answer questions about the images displayed in addition to the sustained attention task. Introducing this secondary objective created a divided attention situation where the test subject was incapable of providing their full attention to a single screen. A visual task was chosen to stay in the same modality as the visual continuous performance task so that both tasks could place demands on the same sensory resources (eyes). This provided an oportunity to detect when the subject's focus was away from the sustained attention task and determine its impact on test performance.

In addition to postural features, this test also used the Kinect to record noise levels

Delay	Count	Median RT	Mean RT	SD RT
1000	13	742	909.923	356.629
2000	16	1519.5	5484.38	15376.4
3000	4	1390.5	1317.25	561.523
4000	10	832	969.5	391.935
5000	18	804	1009.39	571.046
6000	14	881.5	2847.36	6920.1
7000	16	925	963.625	415.788
8000	15	619	882.867	510.153
9000	15	1004	1510.07	1717.56

Too Fast:	0
Correct:	10
Lapse:	110
Sleep Attack:	1

Table 4.6: Example PPVT report statistics.

in the room during the exam. This provided the opportunity to assess the effect of distractions in modalities other than the primary objective of the test subject. The primary noise recorded is that of the subject or the test administrator talking. Subjects were not required to talk or answer questions during the exams, but nore were discouraged from doing so. This provided a range of volume levels for comparison.

Test	Correct Targets	Omissions	Commissions	Mean RT	SD RT	Correct RT Mean	Error RT Mean	Anticipations	Number of Lapses	Target
P-CPT	x	x	x	x	x	x	x			non-X letters
TOAV	x	x	x	x	x	x	x	x		geometric shape
PPVT				x	x				x	geometric shape

Table 4.7: Description of the features measured by each task in the test battery.

Group Id	PCPT	PCPT(DA)	TOAV	PPVT
Group 1	x		x	x
Group 2		x	x	x

Table 4.8: Summary of administered tests by group.

4.3.2 Kinect

The Microsoft Kinect sensor was used during the battery of sustained attention tasks to collect posture and movement data. This data is categorized into two groups, cross-sectional and aggregate features. The Kinect was selected because of its ability to process and translate depth information into articulated figures within a defined three-dimensional space. This greatly simplifies the effort needed to track different parts of an individuals body over time.

Cross-sectional Features

Cross-sectional features are collected throughout the exam and were correlated with the subject's response on a per target basis. That is, for each target/response pair the value of the cross-sectional features was examined. The study's hypothesis was that a pattern will emerge in which subjects respond correctly to a target but have a higher than average response time p, % of the time they present feature x. The following cross-sectional features were assessed and recorded:

Distance Head

Proximity of head from display screen. Calculated as the distance (in millime-

ters) between the reported head joint and the location of the Kinect sensor.

Body Orientation

Measures the difference between the left and right shoulder depth values (in millimeters). This provides an estimation of the direction the body is facing. If the shoulders are square with the computer this value will be close to zero. As the body rotates away from the computer, this value increases.

Head Pose

Heading (in degrees) of the subjects face relative to the location of Kinect sensor. Directly facing the sensor is treated as a heading of zero degrees. This is calculated for yaw, pitch, and roll or rotation of the face around the y, x, and z axis, respectively.

Hand Position

Position of each hand in relation to the Kinect sensor. Measured in X,Y, and Z coordinate space with the Kinect sensor at 0, 0, 0.

Forward Body Lean

Angle and direction of back lean in relation to the hip. This is the angle between the Y component of the Hip_Center-Shoulder_Center edge and the Kinect Y axis.

Side Body Lean

Angle and direction of back lean in relation to hip. This is the angle between the X component of the Hip_Center-Shoulder_Center edge and the Kinect X axis.

Talking

Measures whether the individual is talking during any particular moment in time. Talking is a good measure of inattention in situations where the subject is supposed to be attending the stimulus but not verbally responding to it. This generally indicates there is a distraction and that the subject is at least partially attending to that over the target stimulus.

Aggregate Features

Aggregate features were calculated at the end of the exam and were correlated with the subjects' overall performance compared to the average. In this study we are looking to find an association between trends and overall performance. The study's

hypothesis was that individuals with a greater number of head movements would be identified as more inattentive than their peers by the collected CPT results. The aggregate features used in this study were:

`Distance Head Mean` Average value of the Distance Head feature during the recorded time period.

`Distance Head Delta` Represents the variance in Head joint positions. Provides a measure of how far the head was moving during the designated time window.

`Body Orientation` Average values for each metric reported in the body orientation.

`Body Orientation Delta` Represents the variance for each metric in the body orientation.

`Gaze Percentage` Percent of time gaze was directed at target.

`Movement Head` Total amount of movement by the Head vertex during the recorded time window. Calculated as a summation of the movement recorded for between each frame.

`Forward Body Lean Mean` Average value of the Forward Body Lean feature during the recorded time period.

`Forward Body Lean Delta` Represents the variance in Forward Body Lean angles.

`Side Body Lean Mean` Average value of the Side Body Lean feature during the recorded time period.

`Side Body Lean Delta` Represents the variance in Side Body Lean angles.

`Talking Percentage` Percent of time the individual spent talking during the exam.

Raw Measures

Kinect Studio Along with the Kinect SDK, Microsoft has released a suite of software examples and utilities. Amongst these is an application called Kinect Studio. This application allows a person to record the entire depth and video stream from a Kinect so that it can be played back at a later point in time. Kinect Studio was used in this application to fully record each test subject during each attention task. This recording was later run through several applications to extract detailed feature data at a high rate (30 frames per second). The features recorded from

these applications were written to comma separated value (csv) files to be processed further for analysis.

Joint Data: The CSV file for joint information contains depth and x,y,z values for 12 of the 20 joints the Kinect is capable of tracking. Since we are working with seated test subjects whose legs are partially occluded behind a table, it was difficult to track any joints below the hip.

Head Pose Data: The CSV file for head pose and gaze tracking data contains recorded values for head yaw, pitch, and roll.

Audio Data: The CSV file for audio contains a calculation of volume, source angle, beam angle and confidence. The beam angle describes the angle of the Kinect audio array. This can be adjusted at 10-degree increments between -50 and 50 to point toward target of interest. Accurately pointing the beam angle of the Kinect helps to improve the quality of audio collected. The reported source angle is an estimation of the direction the strongest speech is coming from. The source angle is accompanied by a confidence value, between 0 and 1, describing how reliable the reported source angle is.

Application	Contents	Test	Ouput Filename
KinectAudioRecorder	Timestamp, Confidence, Source Angle, Beam Angle, Volume	pcpt toav ppvt	subject.pcpt.audio.csv subject.toav.audio.csv subject.ppvt.audio.csv
KinectJointRecorder	Timestamp, All Joint Depths, All Joint Positions (X,Y,Z)	pcpt toav ppvt	subject.pcpt.joint.csv subject.toav.joint.csv subject.ppvt.joint.csv
KinectGazeRecorder	Timestamp, Yaw, Pitch, Roll	pcpt toav ppvt	subject.pcpt.gaze.csv subject.toav.gaze.csv subject.ppvt.gaze.csv

Table 4.9: Table of CSV output files for each feature extraction application.

Body Lean

Body lean is the most commonly discussed posture for determining engagement [5] [19] [18]. One objective of this research was to determine if there is a relationship between forward body lean and attention. We know that users who are engaged tend to lean forward more than those who are not. Hence, identifying that a person is leaning forward is a strong indication that the person is also paying close attention. Leveraging the Kinect skeleton tracking framework forward body lean will be calculated as the difference between the hip joint and head Z value. Large positive values for body lean indicate the subject is leaning toward the screen, values close to zero indicate an upright posture, and negative values indicate the subject is leaning back (away from the screen).

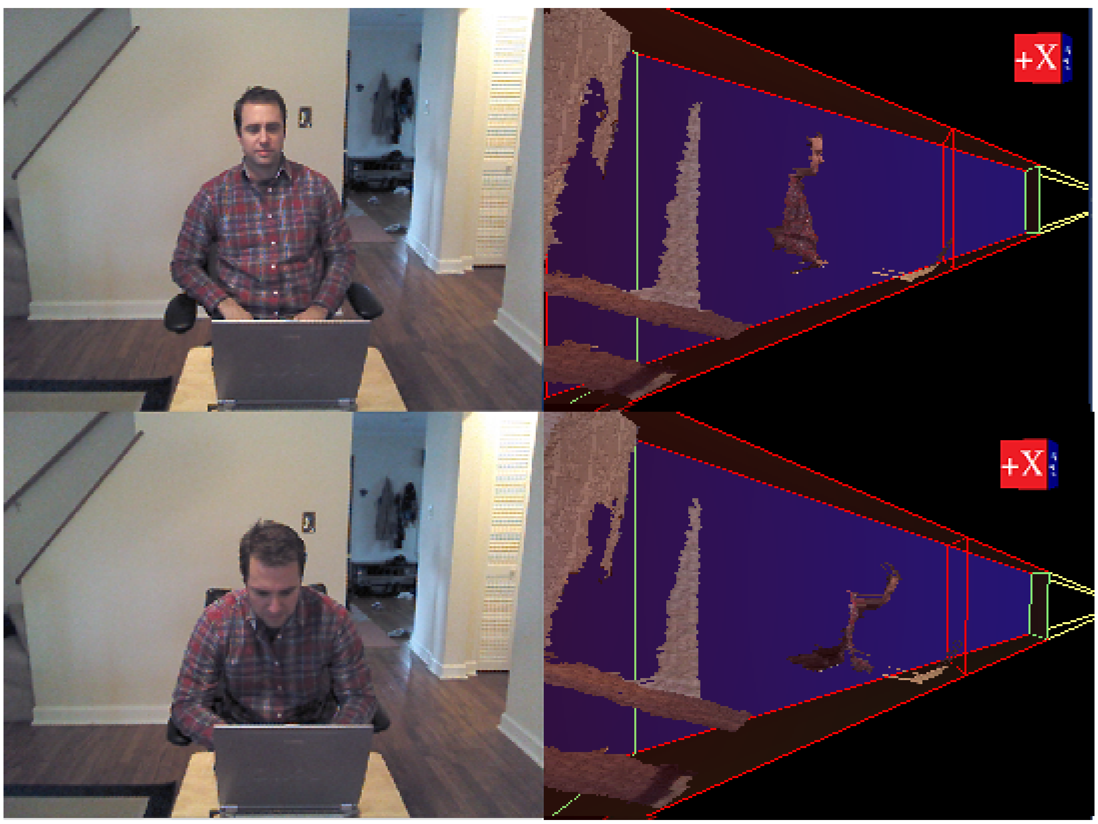


Figure 4.1: RGB and 3D views from Kinect studio comparing test subject lean posture before and during the attention task.

Side Lean

Side lean is a less explored measure of attention and engagement than forward body lean. It was also be examined during this experiment to see if an upright posture is correlated with attention. Side lean is expected to be similar to measurements of slouch discussed by [19], and can be visualized as a person slouching sideways rather than forward. This measurement is calculated as the different between the hip joint and head position X value.

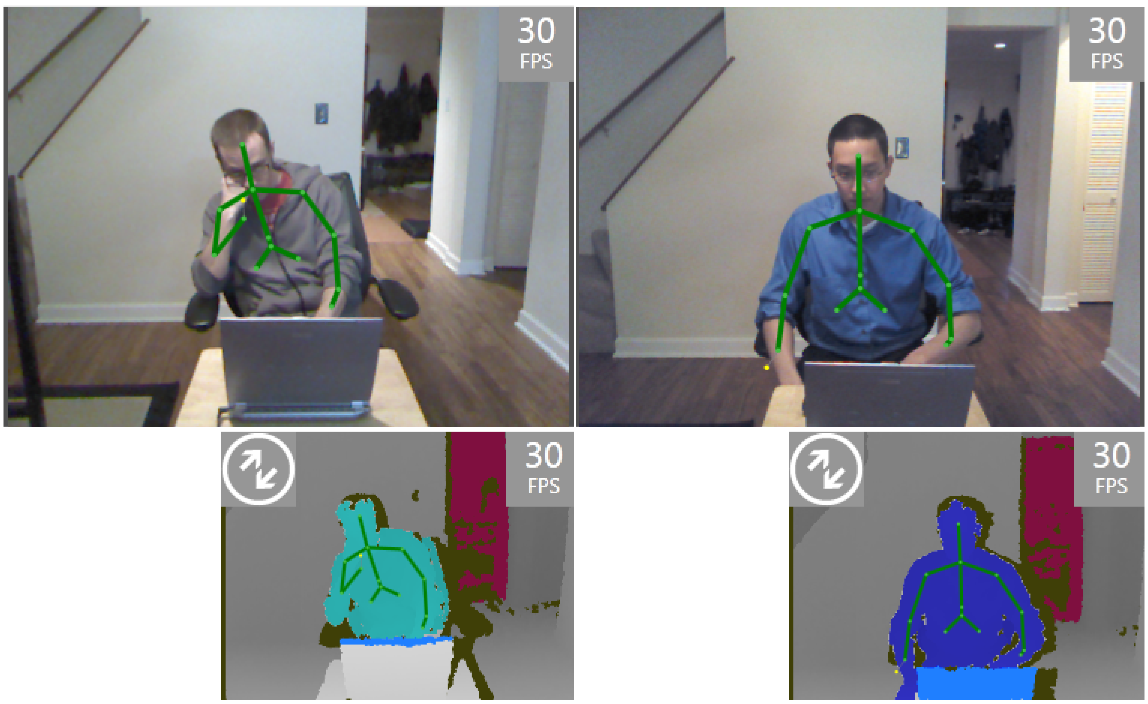


Figure 4.2: RGB and depth views from joint recording application comparing side body lean.

Body Orientation

Body orientation is a measurement of whether a persons upper body is directed at the computer screen. This is a measurement than can be used to determine if the user is facing the computer screen in an upright and proper fashion. Body orientation is calculated as the difference between the left and right shoulder position Z values.

Calculated values start close to zero with the shoulders square to the computer and increase as the body turns away. The further the orientation measure is from zero the less attentive the person is assumed to be.

Head Depth and Movement

Head depth is closely related to Body Lean, but looks solely at the proximity of the head to the screen and is not concerned with the body's angle of inclination. It is possible to be far from the computer screen yet still have a significant Body Lean. Movement is a features discussed by [5] and [12] as correlating to attention and engagment. This study only examined movement of the head since most postural shifts change the location of the head. Hands and feet are also useful measures of body motion but both are largely occluded in the current setup and therefore not utilized as a measure in this study.

Head Pose (Gaze Tracking)

As of 1.5 the Kinect SDK offers face-tracking abilities which this study leveraged to estimate gaze direction. The Kinect uses both the depth and RGB data streams to build and track an 87 point model of the human face. Along with the 87 facial points identified and tracked by the Kinect, a calculation of the subjects head pose is provided. Head pose values of pitch, roll, and yaw can each be tracked between -90 to 90 degrees.

This study used the provided measurement of head pose to estimate the subjects gaze direction. A similar estimation of gaze was used by [2] to determine which gaze target students were looking at based on their head pose. This estimation is based on the theory that subjects frequently turn their head in the direction they are attending. [2] supported this theory with research from Stiefelhagen et al [20] and others who found eye gaze to correlate with focus of attention and head pose to commonly agree with eye gaze.

For the purpose of this study, subjects were classified as looking at the target if each head pose angle (yaw, pitch, and roll) was less than a predetermined threshold. The subject looking directly at the Kinect sensor register head pose values of (0.0, 0.0, 0.0) for yaw, pitch, and roll respectively.

A more accurate assesment of gaze would be to use Eye Gaze detection; however, at

this time eye gaze can be difficult to calculate without expensive equipment, such as high resolution cameras. [15] has done some interesting work on a technique which utilizes the Kinect face tracking in combination with several image processing techniques to obtain an estimation of eye gaze more accurate than is traditionally possible from low resolution cameras such as the one installed in the Kinect. This could be applied to future versions of this software to obtain better gaze direction estimations.

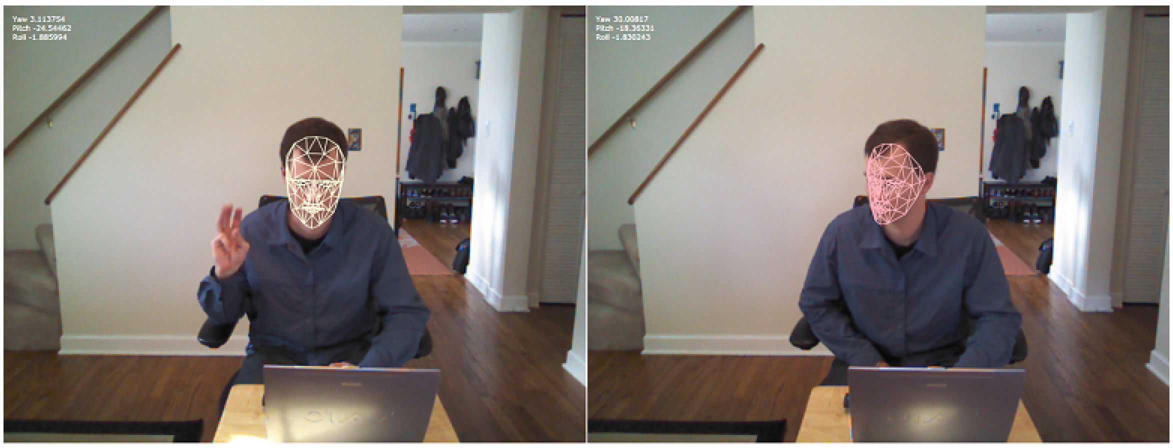


Figure 4.3: Comparison of from head pose tracking of subject looking at and away from screen.

Audio

In addition to body posture and head pose, this system measured the amount of noise in the room during the attention tests. Utilizing the Kinect's build in microphone, volume and source direction were tracked and used to calculate an estimation of when the test subject was talking. Audio tracking is used to get an idea of how much interference from a different modality will affect a subject's attention.

4.4 System

The data collection system is a pipeline of applications exporting summary data files for further processing. The first application used is PEBL, to run the attention test battery and collect measures of attention. This was the only application with which

the test subject interacted during the experiment. Each subject performed three sustained attention tasks (outlined above) within the PEBL application. The result of each task was written to a Comma Separated File (CSV) which captured the user id, time of each test stimuli presentation, and all of the attention measurements specific to that task. While the attention task was being conducted, the Kinect Studio application (provided by Microsoft) was being run on a separate machine connected to a Kinect sensor located on a tripod directly behind the PEBL test computer and facing the user. Kinect Studio was used for recording time-stamped RGB, depth, and articulated data (at a rate of 30 frames per second) to an external file.

The second phase of the pipeline is a series of applications responsible for extracting a particular set of data from the saved Kinect Studio File. The applications are titled: AudioRecorder, JointRecorder, and GazeRecorder. These save audio, articulated figure, and head pose data to another set of CSV files.

The last stage of the pipeline is an application responsible for combining the results from both sets of CSV files into a summary file containing both measures of attention and posture. A time window was used to select which values from the recorded Kinect data files were associated with results of the attention tasks. This allowed for an analysis of data within each test subject by asking summary questions about the values of a feature before and after the presentation of each target during a PEBL attention task.

4.4.1 Hardware Architecture

Participants were asked to take the attention test battery during a single session, lasting approximately one hour. During testing, the subject was seated in a traditional desk chair, including rolling feet and arms with a small notebook computer placed on a desk directly in front of them. The laptops built in keyboard and mouse were used for input during the attention task. The Kinect sensor was mounted 3 inches above the computer monitor approximately 3 feet from the student and facing directly at them. The exact distance of the Kinect from the student varied based on how close the student chose to sit from the screen. This configuration is believed to be a realistic estimation of how subjects might engage in automated tutoring systems or remote learning classrooms which is the main focus of this study.

Kinect Setup

The Kinect for Windows sensor was placed directly in front of the test subject on a video camera tripod behind the test machine. For the first round of testing the following configuration was used.

Distance	36” (from computer); 56” (from test subject)
Height	49”
Sensor Angle	-8 degrees (approx)
Desk Height	26”

This architecture allowed for the collection of 1) Distance Head 2) Distance Torso 3) Body Orientation 4) Gaze Direction 5) Head Position 6) Hand Position 7) Volume Of Motion and 8) Side Body Lean.

4.4.2 Software Architecture

This study utilized software applications freely available from Microsoft and PEBL in addition to software developed solely for the purpose of the experiment. This section describes the role of each software application in the study.

Third Party Software

PEBL - The Psychology Experiment Building Language (PEBL) is an open source software project licensed under the GPL and designed to provide access to basic implementations of well known psychological experiments. This study utilized three of the provided experiments, the PCPT, TOAV, and PPVT.

Kinect Studio - Kinect Studio is a software application developed and distributed by Microsoft alongside their Kinect SDK. The Kinect Studio allows for the recording and playback of entire RGB and depth sensor streams. At this time the Kinect Studio does not support recording or playback of audio and is not distributed along with its source code. Additionally, the format used to store sensor data is currently unpublished. Recorded data can only be accessed by connecting into an attached Kinect and simulating a live stream. For this experiment the Kinect Studio was used to record the RGB and depth stream for later processing in a series of data extraction applications.

Developed Software

KinectAudioRecorder [C#] - This application was developed to record descriptive details about noise detected by the Kinect during the exam. The application is written in C# and is a modified version of the Microsoft Kinect SDK's KinectExplorer application. This application was run during the exam and provided the initial Kinect data stream for the Kinect Studio application to connect to.

KinectJointRecorder [C#] - This application was developed to record joint position and depth values from the Kinects tracked skeltal frame. The application is a modified version of the Microsoft Kinect SDK KinectExplorer example and is run after the initial recording of the depth and RGB data streams by the Kinect Studio application.

KinectGazeRecorder [C#] - This application was developed to record head pose information for estimating gaze direction. The application is a modified version of the Microsoft Kinect SDK FaceTrackingBasics example and is run after the initial recording of the depth and RGB data streams by the Kinect Studio application.

KinectFeatureBuilder [Java] - The KinectFeatureBuilder application was written in Java and was utilized to extract higher level features from the raw audio, joint, and gaze data files, including side and forward lean, which are interpretations of the raw joint position data. This application is also used to aggregate data over a time window around the event, similar to a moving average, to analyze postural features or changes leading up to an event.

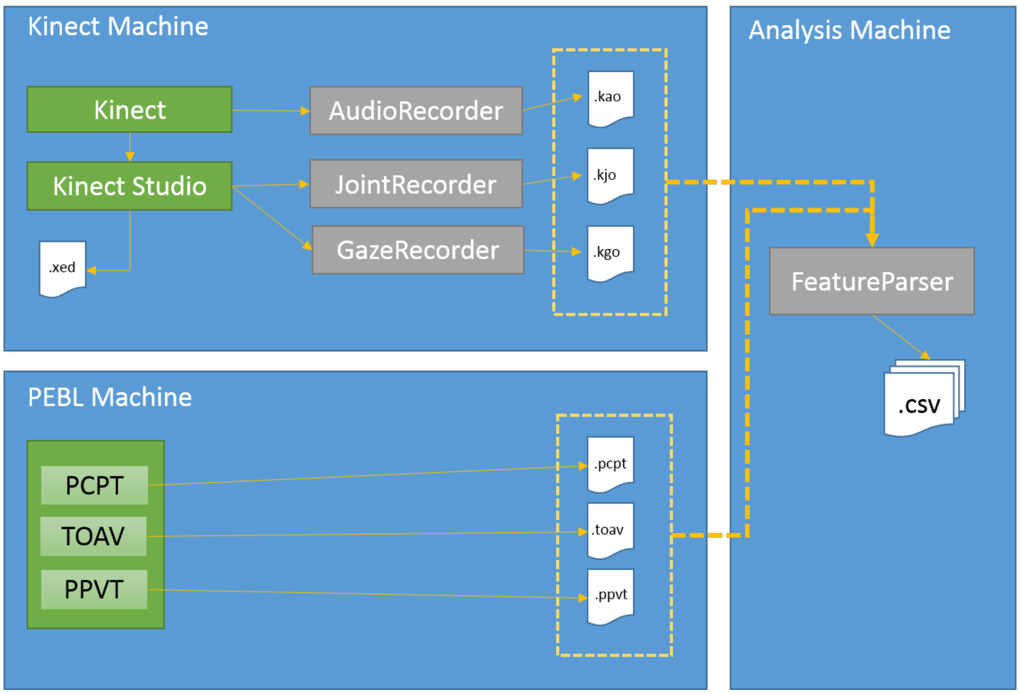


Figure 4.4: Overview of software architecture. Systems in gray were developed for the purpose of this research. The end results of the application pipeline is a series of stacked CSV files with subject features appropriate for data analysis using applications such as Minitab, Excel, and DataDesk.

Chapter 5

Analysis

The data collected during this experiment contains information about each test subject for each test taken. This analysis looks at patterns of behavior 'within' and 'between' test subjects.

Within:
: Within analysis refers to analysis of data within a single test subject, or analysis of collected features during a single test execution, (cross-sectional).

Between:
: Between analysis looks at the differences between test subjects in an attempt to find correlations across test subjects attention and body posture using the aggregate data collected from the Kinect and sustained attention tasks.

How Results Are Interpreted Most sustained attention tasks, especially CPTs, use a combination of recorded responses to determine levels of attention. The most basic of these measurements is the reaction time for each response, number of commission errors and number of omission errors. To determine if a subject displays traits of inattention, the average reaction time, reaction time variance, and number of errors of omission and commission are compared to normal values for the test subjects age and sex. If the subject falls outside of these ranges it is generally an indication of inattention. Often these values are not analysed independently. If a subject has a slower than normal reaction time, but a better than normal omission/commission score, it can be interpreted as an indication that the subject is more deliberate than his peers. This is important to keep in mind since the relationship

between these different values can be overlooked during single variable correlation analysis.

5.1 Body Posture

This section will examine the relationships between posture and attention both within and between test subjects.

5.1.1 Within

The first set of results were the measures of body posture compared to attention within individual test subjects. Reaction time (RT) and reaction time standard deviation (RT STD) were independently treated as measures of attention for this analysis. For RT, smaller RT values indicated better attention. Likewise, smaller RT STD values are indicative of better attentiveness.

A linear regression was calculated, first with RT as the response variable and then with RT STD as the response variable for each of our cross-sectional measures of posture, HD, BODY LEAN, and SIDE LEAN, as well as their standard deviations. A resulting P-value smaller than 0.05 is considered significant and indicates a possible correlation between the feature and the response. Linear regression results for this calculation did not result in significant or consistent correlations for any of the proposed measures of posture and reaction time.

One explanation for the lack of a correlation between measures of posture and attention is that a lack of sufficient body posture variance during a single attention exam. The lack of variation in body posture for this sample is confirmed by a quick visual analysis of the tested features. 5.1 shows a line plot of head depth over time, during the course of the TOAV exam. Similar visual analysis was conducted for each feature and also across the PCPT and PPVT exams. Each inspection showed similar results. This confirmation that our tested features vary infrequently during the course of the exam is an indication that analysis within test subjects might not be appropriate for this data set. A notable exception to this observation, was one test subject whose head depth changed significantly during the exam. This exception is discussed further during case studies in section 5.6.

From the results of the linear regression, along with a visual inspection of the data

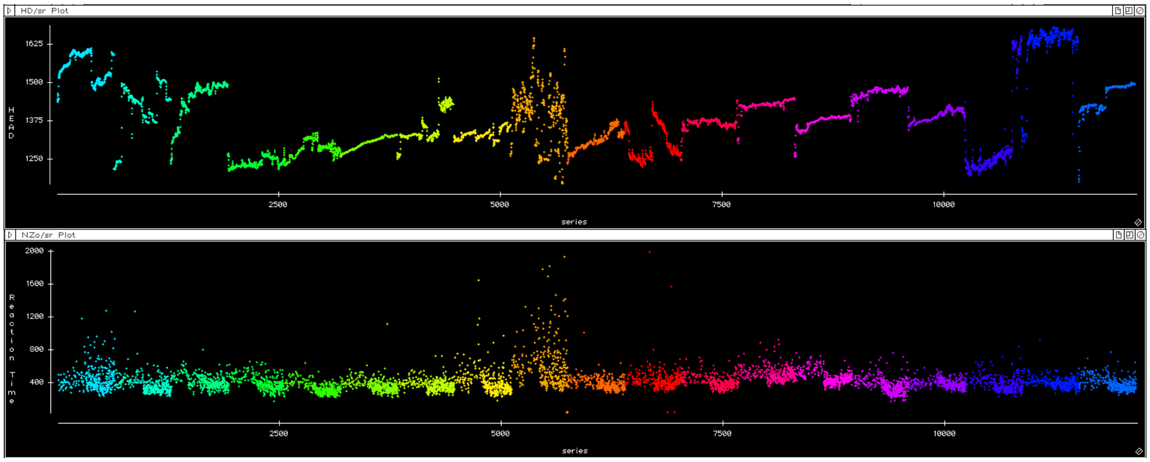


Figure 5.1: A side by side comparison of Head Depth and Reaction Time. Each unique color represents a different test subject.

it is clear that, using reaction time and omission rate, most test subjects were classified as attentive. Considering our hypothesis that movement is an indication of of inattention, the observed lack of motion during the exams is not a surprising result.

5.1.2 Between

This analysis looked at the measured attention and posture differences between test subjects in attempt to find correlations between test subjects attention and body posture during the test to determine if test subjects with particular body posture behaviors tend to perform better on attention tasks. We started by looking at the body posture feature of head depth. Given that engaged subjects tend to lean forward, bringing their head closer to the screen, we expected to find that test subjects with smaller average head depth during the course of the exam had better average reaction times. A visual analysis confirmed that several of the test subjects who sat closest to the screen performed well in regard to average reaction time.

A linear regression was calculated, first with RT as the response variable and then with RT STD as the response variable for each of our aggregate measures of posture, Head Depth (HD), Body Lean (BLEAN), Side Lean (SLEAN), Body Orientation (SDIFF), and Movement (MOVE) as well as their standard deviations. A resulting P-value smaller than 0.05 is considered significant and indicates a possible correlation

between the feature and the response.

Linear regression results for the relationship between BLEAN and RT were statistically significant for the PCPT and combined regression tests; this was also true for BLEAN and RT STD. For the TOAV exam, measures of Body Orientation (SDIFF) and Movement (MOVE) were statistically significant for both RT and RT STD. Interestingly, Head Depth showed a relationship with RT only for the combined results. This is possibly related to the fact that test subjects expressed greater interest in, and performed better on, the PPVT exam and therefore tended to sit closer to the screen. A comparison of the collected results across all tests reflects this relationship.

One explanation for why the relationship between Movement and RT was only identified during TOAV is that, on average, test subjects were sitting very still during the attention exams. The length of the TOAV (24 minutes) may have caused greater discomfort or boredom with test subjects, leading toward more restless behavior.

A full description of the linear regression results for posture features are described in 5.1 and 5.2.

	HD	HD STD	BLEAN	BLEAN STD	SLEAN	SLEAN STD	SDIFF	SDIFF STD	MOVE	MOVE STD
PCPT	0.13	0.877	0.056	0.833	0.333	0.99	0.575	0.589	0.759	0.963
TOAV	0.385	0.601	0.261	0.879	0.52	0.49	0.403	0.457	0.004	0.013
PPVT	0.263	0.862	0.24	0.824	0.314	0.658	0.145	0.606	0.431	0.719
Combined	0.008	0.933	0.01	0.845	0.091	0.431	0.1	0.789	0.443	0.331

Table 5.1: Single variable regression P-Values for posture features using Reaction Time as the response. P-Values smaller than 0.05 are considered statistically significant and are highlighted in gray.

	HD	HD STD	BLEAN	BLEAN STD	SLEAN	SLEAN STD	SDIFF	SDIFF STD	MOVE	MOVE STD
PCPT	0.073	0.856	0.01	0.905	0.262	0.585	0.632	0.835	0.75	0.861
TOAV	0.562	0.381	0.391	0.414	0.097	0.026	0.019	0.037	0.004	0.041
PPVT	0.449	0.972	0.38	0.386	0.612	0.72	0.624	0.767	0.381	0.524
Combined	0.08	0.856	0.031	0.981	0.258	0.341	0.243	0.441	0.988	0.822

Table 5.2: Single variable regression P-Values for posture features using Reaction Time Standard Deviation as the response. P-Values smaller than 0.05 are considered statistically significant and are highlighted in gray.

The results of the regression analysis on body posture features was somewhat mixed. For our within subjects analyses only a few subjects showed a relationship between RT and measure posture features, such as Head Depth. The regression results between test subjects, in contrast, showed both a relationship between Body Lean and RT, and a relationship between Movement and RT.

5.2 Head Pose

For visual attention tasks, such as the those used for this experiment, eye gaze should be a clear indication of whether the test subject is attending to the target. Here we review the usage of head pose, obtained from Kinect face tracking, for gaze estimation. For the purpose of this analysis, the following metrics were defined for head pose:

Look Away (LA)
: A measurement signifying when the subject looks away from the screen. The subject is considered to be looking away from the screen any time the YAW value of the head pose is greater than 15 degrees away from 0.

LA COUNT
: The number of recorded frames where the subject is determined to be looking away from the screen.

LA OM
: The number of omission errors that occurred while the YAW was greater than 15 degrees away from 0.

LA OM %
: The percent of omission errors that occurred while the subject was classified as looking away from the screen.

The strongest indication of in-attention in a CPT is the number of omission errors. Overall, our test setup witnessed few omission errors as was expected since the test group consisted of normal adults. The division of the PCPT exam into two groups with Group 2 participating in a divided visual attention task provided an excellent distraction which was easily measurable. A comparison of the number of omission between Group 1 and Group 2 shows a clear distinction between the two groups. Subjects participating in the divided attention version of the PCPT averaged 39.90 omission errors while subjects participating in the traditional PCPT exam only average 1.89 omission errors.

Overall, the system proved capable of face tracking which provided reliable head pose calculations. The system did, however, have some trouble tracking tall test subjects (over `6'3"`) due to the height of the computer screen relative to the test subject. Taller subjects had a higher downward(pitch) head angle which decreased the systems ability to recognize their face. The system especially struggled with two of the test subjects from Group 2, subject 2 and subject 18. The calculated look-

away metrics for these subjects reflected the system's inability to accurately assess head pose. These cases are highlighted in Table 5.3 and clearly stand out as the only look-away metrics which did not correspond to the increased omission errors for the Group 2 test subjects.

The PCPT data set shows that 44.10% of omission errors occurred when subjects had a YAW value greater than 15 degrees. If we remove the suspect data from test subjects 2 and 18 we get 76.02% of omission errors occurring with YAW greater than 15 degrees. It is important to note that of all the measured YAW values only 12.84% of them were greater than 15 degrees.

SUBJECT ID	GROUP ID	OM	CM	RT	RT STD	LA COUNT	LA OM	LA OM %
1	2	53	11	474.80	387.44	9789	41	0.77
2	2	59	14	460.02	320.87	218	0	0.00
3	2	30	22	348.05	142.68	3610	24	0.80
4	1	3	10	318.83	76.31	490	3	1.00
5	2	19	22	342.35	91.14	5199	15	0.79
6	1	0	13	318.69	79.14	0	0	0.00
7	1	1	13	329.11	50.37	17	0	0.00
8	1	7	13	341.64	78.80	0	3	0.43
9	2	20	32	371.57	184.59	3992	13	0.65
10	1	1	8	447.68	238.27	296	1	1.00
12	1	2	15	355.23	82.30	0	1	0.50
13	1	0	13	335.49	75.15	165	0	0.00
14	2	41	13	433.06	209.37	762	32	0.78
15	2	5	6	562.23	214.67	6098	5	1.00
16	2	29	25	438.79	295.79	8087	21	0.72
17	2	24	13	322.16	87.48	3823	19	0.79
18	2	119	14	348.36	107.45	91	0	0.00
19	1	2	11	369.78	87.93	110	0	0.00
20	1	1	17	338.36	81.40	545	1	1.00
Group 1 Avg	1	1.89	12.56	351.54	92.81	172.9	1.7	0.44
Group 2 Avg	2	39.90	17.20	410.14	204.15	4166.9	17	0.63
Group 2 Mod Avg	2						21.25	0.79

Table 5.3: Summary of test subject omission rates for PCPT test. Highlighted in gray are suspect values related to a known deficiency in the system. Group 2 Mod Avg is a calculation of group 2 averages with these values removed.

5.3 Audio

To calculate when a test subject was talking, the volume level and audio source angle were measured. A linear regression was calculated using reaction time and omission

rate as dependent variables to determine if the total amount of time a subject spent talking during the exam affected their attention. Results showed that talking was not a significant predictor of reaction time or omission rate for any of the adminstered tests. This is a somewhat surprising result, but may be explained by the fact that the given attention tasks were entirely visual and required no responses to auditory stimuli. Additionally, a regression analysis was peformed with the total noise in the room as the independent variable to determine if more noise in the room acted as a distraction and lowered the test subjects attention. This regression also showed there was not a significant correlation between noise in the room and attention. This could be because the only noise in the room came from a conversation between the test subject and the test administrator and was controlled by the test subject. This conversation was only maintained while the test subject was asking questions or telling a story. That is, the subject may have been engaging in the conversation when the test was placing relatively few demands on his or her attentional capacity and hence the subject talking did not result in diminished performance.

Past divided attention research shows that individuals are significantly better at multi-tasking across multiple sensory modalities than within the same sense. This supports the non-significant regression results for the auditory features measured. Extending the results obtained from the Head Pose analysis, where multiple stimuli were competing for the same sensory organ, it follows that if given an auditory attention task subject talking and external noise measures would have a larger relationship with attention.

5.4 Multivariate Regression

The analysis thus far has attempted to find correlations between two variables using standard linear regression techniques. While visual inspection of the data seemed to indicate relationships between attention and posture, single variable regression analyses did not attain statistical significance. Consequently, this section explores several other techniques starting with a multivariate regression.

A multivariate regression was run in Minitab using several different sets of predictors. For each regression, reaction time was used as the response variable and the predictors were selected from a pool of posture, head pose, and audio features. The predictors were selected by choosing those with the smallest p-values from the single variable correlation. This was performed independently for each of the three attention tests.

The complete results can be found in Appendix G, while the most significant findings are discussed below.

PCPT Based on single variable regression results, the features chosen as predictors for the PCPT multivariate regression were: AVGHD, HDGROUP, BODYLEAN, YAW, LOOK_AWAY_SUM, and PITCH.

Predictor	Coef	SE Coef	T	P
Constant	-285.1	568.6	-0.50	0.630
AVGHD	0.2788	0.4065	0.69	0.512
hdgroup	-33.67	45.31	-0.74	0.479
TALK%	593.88	96.01	6.19	0.000
BODYLEAN	-0.9020	0.4656	-1.94	0.089
fleangroup	78.19	30.83	2.54	0.035
YAW	0.396	1.993	0.20	0.848
YAWD	-10.666	3.485	-3.06	0.016
LOOK_AWAY_SUM	0.033427	0.007586	4.41	0.002
PITCH	0.478	1.727	0.28	0.789
ROLLSD	19.043	6.150	3.10	0.015

```
S = 39.5593   R-Sq = 92.9%   R-Sq(adj) = 84.1%

Analysis of Variance
```

Source	DF	SS	MS	F	P
Regression	10	164906	16491	10.54	0.001
Residual Error	8	12519	1565		
Total	18	177425			

The R-Sq and P-Value for this regression are strong, showing that the selected set of features is a good predictor of RT STD for the PCPT exam. Reaction time standard deviation is a measure of how much variance was witnessed in test subjects' reaction times, where we expect less attentive subjects to have a higher variance. While this model shows a good overall R-Sq value, it has several high P values, such as Yaw, which means they are contributing little to the overall regression. Additionally, there are several values that may be inter-related and potentially skewing the results (for example AVGHD and HDGROUP). To verify the validity of this regression, these

values were removed to obtain the following results:

Predictor	Coef	SE Coef	T	P
Constant	91.30	54.37	1.68	0.121
TALK%	594.43	83.20	7.14	0.000
BODYLEAN	-0.9789	0.3498	-2.80	0.017
fleangroup	78.52	25.51	3.08	0.011
YAWD	-10.293	2.926	-3.52	0.005
LOOK_AWAY_SUM	0.031251	0.004522	6.91	0.000
PITCH	0.906	1.296	0.70	0.499
ROLLSD	17.062	4.892	3.49	0.005

S = 34.9136   R-Sq = 92.4%   R-Sq(adj) = 87.6%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	7	164016	23431	19.22	0.000
Residual Error	11	13409	1219		
Total	18	177425			

Further reduction of duplicate measures and poor contributors gives us

Predictor	Coef	SE Coef	T	P
Constant	-37.51	35.26	-1.06	0.304
TALK%	477.74	88.04	5.43	0.000
LOOK_AWAY_SUM	0.018755	0.003709	5.06	0.000
ROLLSD	19.426	5.977	3.25	0.005

S = 47.6061   R-Sq = 80.8%   R-Sq(adj) = 77.0%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	3	143430	47810	21.10	0.000
Residual Error	15	33995	2266		
Total	18	177425			

This demonstrates that if we combine the subject's gaze direction and their verbal activity, we have a good prediction of what their reaction time variance will be. This also tells us that body posture features, such as head depth and lean were not as important as where the test subject was looking. This makes sense for PCPT as it was composed of two groups with the second group being requested to attend to a secondary visual stimulus. These results show that participation in the divided attention task had an effect on subject response time and the Kinect was able to capture sufficient information to identify the scenario.

TOAV For the TOAV test, setting RT as the response and selecting the predictors HDGROUP, LDROOP, LDROOP_SD, RDROOP_SD, YAWD, PITCHSD, and ROLLSD resulted in:

```
Predictor     Coef  SE Coef      T      P
Constant   472.76    69.20   6.83  0.000
hdgroup     33.95    19.85   1.71  0.111
LDROOP      -2410     1083  -2.23  0.044
PITCHSD     3.177    5.132   0.62  0.547
ROLLSD      3.818    7.311   0.52  0.610

S = 44.7602   R-Sq = 63.1%   R-Sq(adj) = 51.8%

Analysis of Variance

Source          DF     SS     MS     F      P
Regression       4  44616  11154  5.57  0.008
Residual Error  13  26045   2003
Total           17  70662
```

PPVT For RT STD the selected predictors were HDGROUP, TALK%, BODYLEAN, LDROOP,

```
Predictor      Coef  SE Coef      T      P
Constant     136.53    31.79   4.29  0.001
hdgroup      -15.74    15.67  -1.00  0.331
TALK%        171.54    77.47   2.21  0.043
BODYLEAN    -0.1570   0.1744  -0.90  0.382
LDROOP       -660.1    536.1  -1.23  0.237
```

S = 34.0259   R-Sq = 42.4%   R-Sq(adj) = 27.0%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	4	12771	3193	2.76	0.067
Residual Error	15	17366	1158		
Total	19	30137			

TOAV+PPVT Last, we look at the combined results of TOAV and PPVT. PCPT was left out since it was determined head pose has such a large impact on that result set. Looking at both TOAV and PPVT together will help us identify features that were predictive across multiple attention tasks.

Predictor	Coef	SE Coef	T	P
Constant	85.67	18.17	4.72	0.000
SDHD	0.6768	0.2866	2.36	0.025
AVGVOL	-2862	1019	-2.81	0.009
TALK%	493.8	131.4	3.76	0.001
BODYLEANSD	-2.9684	0.9083	-3.27	0.003
SIDELEANSD	4383	1378	3.18	0.004
YAWD	-7.452	3.533	-2.11	0.044
ROLLSD	8.359	3.485	2.40	0.023

S = 34.2345   R-Sq = 57.6%   R-Sq(adj) = 47.0%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	7	44627	6375	5.44	0.001
Residual Error	28	32816	1172		
Total	35	77443			

Interestingly, Talk% shows a stronger relationship than AVGVOL, indicating that a subject talking interferes with his or her performance more than someone else in the

room talking.

The summary chart for the multivariate regression analysis displays interesting results in the frequency of certain features in the final regression equation. The most frequent features were ROLL STD and TALK% with BODYLEAN, PITCH, and YAW STD close behind. The most surprising value was how often LDROOP was used in the equation and how infrequently SIDELEAN was used. This may indicate that when a subject slumps sideways in a chair they may only be shifting their shoulders and not their entire body. This body posture could be picked up by an LDROOP, but not the calculation used for SIDELEAN.

TEST	RESPONSE	RSQ	PVAL	AVGHD	HDGROUP	SDHD	HDSTDGROUP	AVGVOL	SDVOL	TALK%	BODYLEAN	FLEANGROUP	BODYLEANSD	SIDLEAN	SIDELEANSD	SDIFF	SDIFFSD	LDROOP	LDROOP_SD	RDROOP	RDROOP_SD	YAW	YAWD	LOOK_AWAY_SUM	PITCH	PITCHSD	ROLL	ROLLSD
PCPT	RT	36.4	0.125	x							x											x			x			
PCPT	RTSTD	92.4	0							x	x	x											x	x	x			x
PCPT	RTSTD	80.8	0							x														x				x
TOAV	RT	63.1	0.008		x													x								x		x
TOAV	RTSTD	60.5	0.012				x			x											x							x
PPVT	RT	48.6	0.164								x					x				x			x		x			x
PPVT	RTSTD	42.4	0.067		x					x	x							x										
TOAV PPVT	RT	27.6	0.063	x								x											x		x			x
TOAV PPVT	RTSTD	28	0.012							x								x										x
TOAV PPVT	RTSTD	61.6	0.001			x		x		x	x		x		x		x						x					x
TOAV PPVT	RTSTD	57.6	0.001			x		x		x			x		x								x					x
				2	2	2	1	2	0	7	5	2	2	0	2	1	1	3	0	1	1	1	5	2	4	1	0	9

Table 5.4: Results of multivariate regression for all attention tasks.

5.5 Groups

There is a certain amount of variance that is bound to occur between different test subjects' reaction time averages, even if they are both attending at their maximum capacity. Due to the fact that our sample set consists exclusively of well attending adults this natural difference may have had an affect on our ability to calculate strong single variable correlations between some of our features and RT. One way to help reduce these personal differences is to look at groups of people based on their measured features. This provides an organized system to classify individuals based on measured posture features for attention prediction.

Groups were identified for HD, BODYLEAN, TALK, and YAW. For each group the average RT and RT STD were compared.

5.5.1 Head Depth

The first grouping used was head depth. Features of head depth and head depth standard deviation were divided into three groups. The boundaries for each group were chosen to provide distinguishable groups as much as possible. Group 0 represents values below the average, Group 1 represents the average, and Group 2 represents above average.

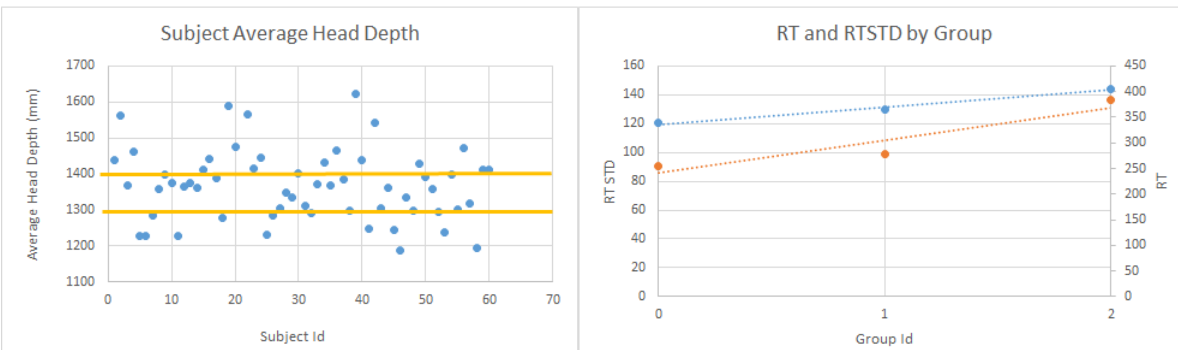


Figure 5.2: Line plot of all test subject average head depths. Horizontal orange bars represent division lines for the three groups, below average, average, and above average.

Group Id	Group Range (mm)	# in Group	Average Head Depth (mm)	Average RT (ms)	Average RT STD (ms)
0	<1300	16	1253.690115	338.6400384	90.75586022
1	1300 - 1400	23	1355.994657	364.3284019	98.55166589
2	>1400	21	1472.020674	405.5424338	136.3229105

Table 5.5: Comparison of RT and RT STD for subjects grouped by head depth.

A brief examination of the groupings for head depth, found in 5.5, shows those grouped with below average head depth values (sitting closer to the screen) had slightly faster average RT and a smaller variance in RT. This group also had an average RT and RT STD which was slower and had more variance. These values appear to agree with the notion that attentive subjects tend to sit closer to the screen.

In order to further explain the statistical significance of these groups an ANOVA was run treating each group id as a factor.

```
Source  DF      SS     MS     F      P
Factor   2   42786  21393  5.51  0.007
Error   57  221383   3884
Total   59  264169
```

The P-Value of 0.007 is a good indication that there is a statistical difference between the RT values of the different groups.

ANOVA for RT STD

```
Source  DF      SS     MS     F      P
Factor   2   23485  11743  2.37  0.103
Error   57  282878   4963
Total   59  306363
```

5.5.2 Head Depth Variance

Features of head depth variance were also divided into three groups. The boundaries for each group was chosen to create as clearly distinguishable groups as possible. Group 0 represents values below the average, Group 1 represents the average, and Group 2 represents above average.

Group Id	Group Range (mm)	# in Group	Average Head Depth Std (mm)	Average RT (ms)	Average RT STD (ms)
0	<40	34	21.75301937	367.2458799	99.81616397
1	40 - 80	15	54.91220139	371.7484665	139.2147832
2	>80	11	107.9527625	386.5089137	99.96289797

Table 5.6: Comparison of RT and RT STD for subjects grouped by Head Depth Std.

ANOVA for RT

```
Source  DF      SS    MS     F      P
Factor   2    3084  1542  0.34  0.716
Error   57  261085  4580
Total   59  264169
```

ANOVA for RT STD

```
Source  DF      SS    MS     F      P
Factor   2   17431  8716  1.72  0.188
Error   57  288932  5069
Total   59  306363
```

5.5.3 Body Lean

Features of body lean were divided into three groups. The boundaries for each group was chosen to create as clearly distinguishable groups as possible. Group 0 represents values below the average, Group 1 represents the average, and Group 2 represents above average.

Group Id	Group Range (mm)	# in Group	Average Body Lean (mm)	Average RT (ms)	Average RT STD (ms)
0	<100	15	62.21773643	385.7450276	133.5500655
1	100 - 150	26	121.8149622	382.4532312	118.5722302
2	>150	19	196.5050335	346.5381864	78.70706491

Table 5.7: Comparison of RT and RT STD for subjects grouped by Body Lean.

ANOVA for RT

```
Source  DF      SS    MS     F      P
Factor   2   17992  8996  2.08  0.134
Error   57  246177  4319
Total   59  264169
```

ANOVA for RT STD

```
Source  DF      SS     MS     F      P
Factor   2   28830  14415  2.96  0.060
Error   57  277533   4869
Total   59  306363
```

5.6 Case Studies

This section takes a closer look at three interesting test subjects. Each subject is referred to by the color used in 5.3 to identify them.

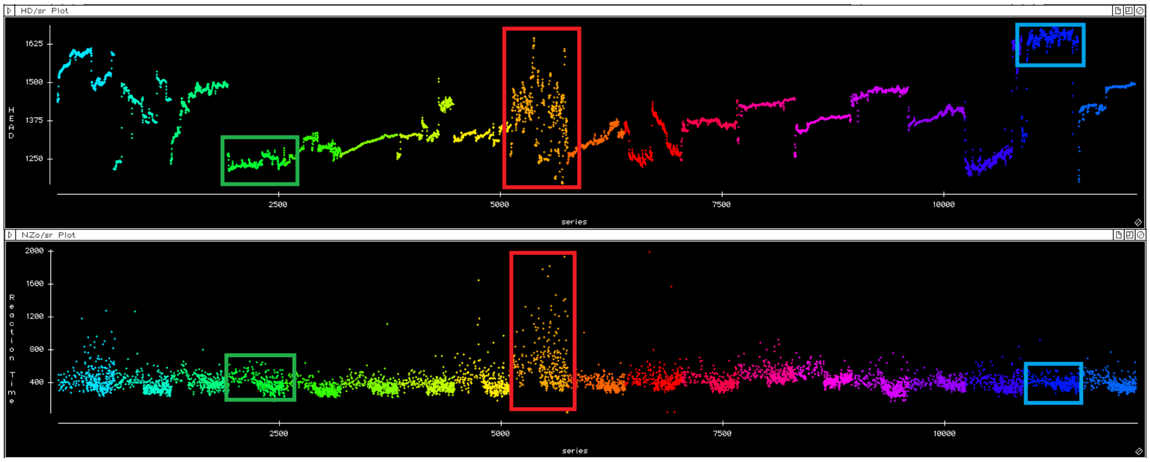


Figure 5.3: A side by side comparison of Head Depth and Reaction Time for all test subjects. Subjects involved in this case study are outlined.

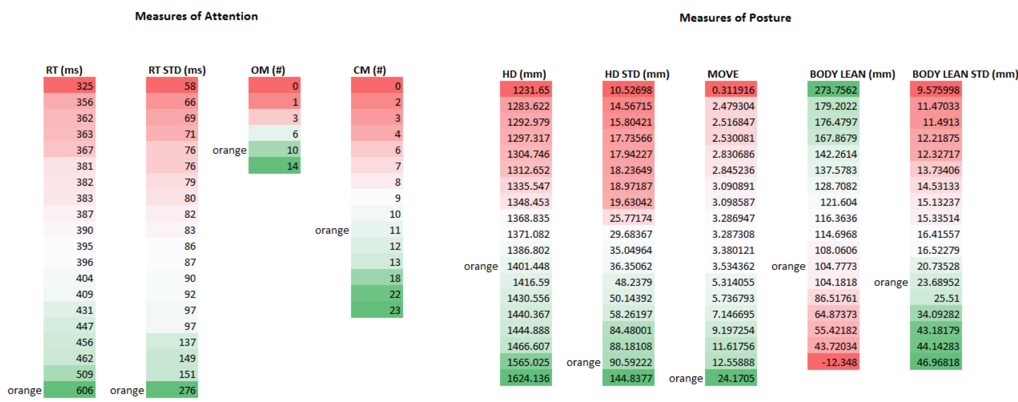
5.6.1 Subject Orange

One of the most interesting test subjects involved in this study, referred to from here out as subject Orange, exhibited postural behavior significantly different from the other test subjects. This was particularly interesting because the test subject's measured attention metrics also stood out from the relatively homogeneous sample. Many of the postural behaviors observed in subject Orange correspond to features discussed in the hypothesis, suggesting that greater variance in attention capacity across the subjects may have helped form stronger correlations.

Let's begin by comparing subject Orange's attention measures with the rest of the group. Figure 5.3 shows there was significantly more variance in RT for Orange than other test subjects, while table 5.8 shows a breakdown of the other attention metrics.

Orange clearly demonstrated the expected behaviors of an inattentive test subject and scored in the bottom 10 percent for all three measures of attention. The slow average reaction time, significant reaction time variance, and high number of omission errors classified this test subject as inattentive. With the knowledge that subject Orange has been classified as inattentive, an analysis of their posture shows positive results. As expected, the inattentive subject appeared in the bottom half for all measured posture features. Most notably, Orange showed a significantly larger head depth standard deviation and a calculated movement score almost twice the next

closest and 5 times the amount of the group average. These results support the stated hypothesis that movement and body lean can be used to predict attentiveness.

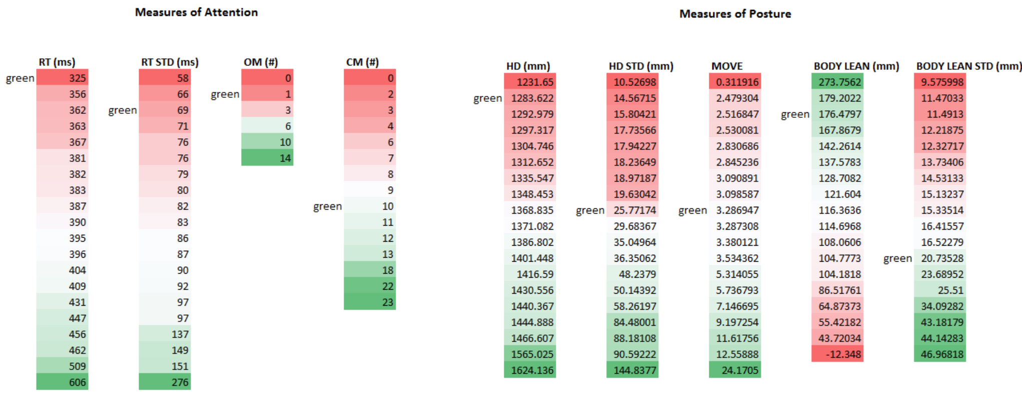


	RT	RT STD	OM	CM	HD	HD STD	BODY LEAN	BODY LEAN STD	MOVE
Orange	606	276	10	11	1401.448	90.59222	104.7773014	23.68952409	24.1705017
Rest Of Group	400.2631579	90.84210526	1.368421053	9	1384.547506	40.80073949	118.1732987	21.37562903	4.7

Table 5.8: Above: chart of subject Orange's position relative to others for each individual measure of attention and posture. Below: Table comparing Orange's attention and posture metrics to the mean of the remaining test group.

5.6.2 Subject Green

Subject Green is a good example of expected behaviors for an attentive test participant. The subject scored in the top 15 percent for both reaction time and reaction time standard deviation, while only committing one omission error. The subject also had a relatively average commission error rate, but this had no impact on the overall attentiveness rating. Green's posture showed they sat close to the screen, with a large amount of forward lean, and with an average amount of body movement. Green exhibited postural traits consistent with research on engaged users, which, combined with Green's attention metrics, support the claim that engaged users are attentive.



	RT	RT STD	OM	CM	HD	HD STD	BODY LEAN	BODY LEAN STD	MOVE
Green	325	69	1	10	1283.622059	25.77173509	176.4796524	20.73527757	3.286946886
Rest Of Group	415.0526316	101.7368421	1.842105263	9.052631579	1391.093377	44.40187747	113.9555134	21.54940824	5.733369444

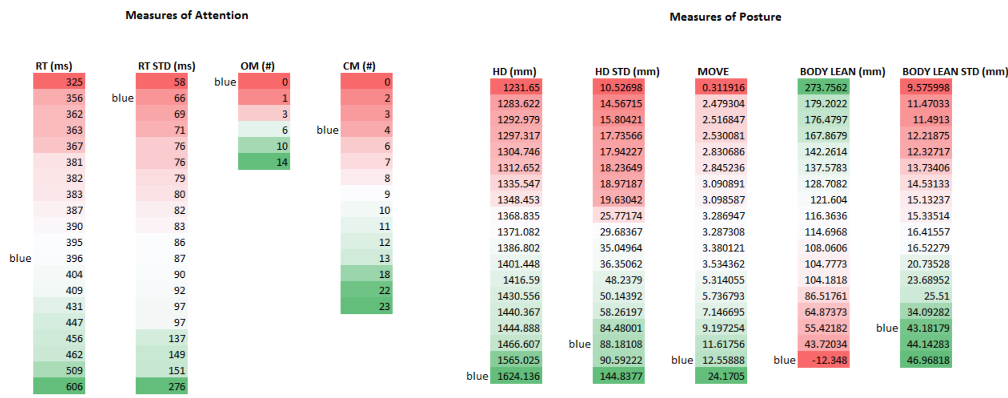
Table 5.9: Above: chart of subject Green's position relative to others for each individual measure of attention and posture. Below: Table comparing Green's attention and posture metrics to the mean of the remaining test group.

5.6.3 Subject Blue

Subject Blue showed up frequently during the various regression analyses as an outlier. Blue is particularly interesting because, unlike Orange and Green, exhibited behavior contrary to this study's hypotheses. Blue performed rather well on the attention task, while scoring in the bottom 15 percent of all measured postural features. They were the only test subject to be measured with a negative average body lean, indicating they were leaning back in the chair rather than forward. This demonstrates that leaning toward the screen does not necessarily provide higher scores on attention tasks. This may instead be an example of a behavioral trait that affects human posture when one is engaged or in a state of readiness; These postural tendencies are not something the subject is consciously thinking about (or doing) in order to improve their attention levels.

Hence, there may not be a physiological reason why a posture such as leaning forward should improve our attention. Blue demonstrates this fact clearly. One explanation of Blue's behavior is the Hawthorne effect. The Hawthorne effect is a phenomenon where observed subjects modify an aspect of their behavior due to the knowledge that they are being observed. During the course of the exam, Blue mentioned to the test administrator that he was surprised that his reaction time scores were no different

when he got closer to the screen. The test subject had not been told the purpose of the experiment, but was aware he was being recorded by a Kinect. The subject's revelation indicates that not only was he consciously aware of his body posture, but that he had also surmised its relationship to given exam. This self awareness could explain why the subject had such a large head depth standard deviation and why he was able to achieve such a high attention score while sitting far away from the screen.



	RT	RT STD	OM	CM	HD	HD STD	BODY LEAN	BODY LEAN STD	MOVE
Blue	396	66	0	4	1624.136169	88.18107808	-12.34799423	43.18178712	12.55888023
Rest Of Group	411.3157895	101.8947368	1.894736842	9.368421053	1372.175926	40.93469175	125.063022	20.22902532	5.3541744

Table 5.10: Above: chart of subject Blue's position relative to others for each individual measure of attention and posture. Below: Table comparing Blue's attention and posture metrics to the mean of the remaining test group.

5.6.4 Test Edges

A visual inspection of subject head depth during the PCPT exam revealed an interesting pattern. Color coded head depth values from before the test began, during the test, and after the test's conclussion showed that before the test began and after the test ended subjects displayed significant movement and sat further from the screen than during the test. This is easily explainable, as test subjects who were leaning toward the screen sat up after the test had ended. This indicates that most test subjects were indeed leaning toward the computer during the attention task.

To determine how significant this finding was, four test subjects were recorded for 5 minutes before the attention task was given. These subjects were not told the

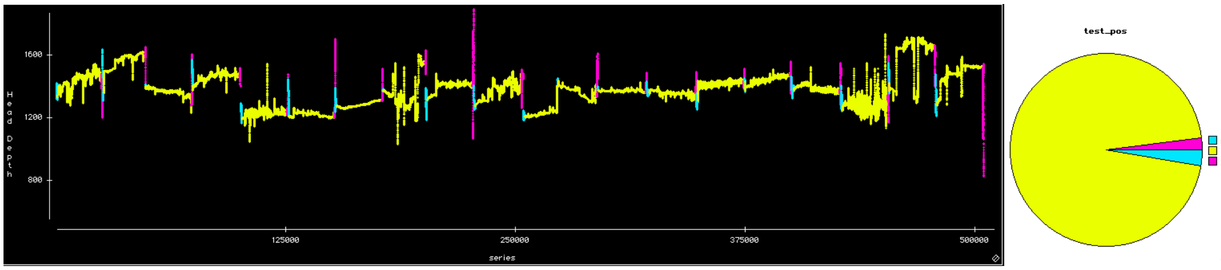


Figure 5.4: Image of head depth for all test subjects over time during the TOAV attention exam. Measured values are grouped and color coded based on on time of occurrence. Blue) Before first target Yellow) After first target Pink) After last target.

Kinect was recording yet and were not required to look at, or pay attention to, the test machine. The goal behind this brief experiment was to determine if there was a difference between the postures of subjects involved in the task compared with those who were not yet engaged with the task. This investigation only looked at the features Head depth (HD) and body lean. Three of the four test subjects displayed a HD significantly larger than the test average. One test subject displayed posture features significantly different from the test average in every category. Due to the small sample size of this experiment it is difficult to draw any statistically significant conclusions; however, in combination with the observed behavior of head depth spikes at test start and end, there is reason to believe test subject posture changes significantly when attention is requested.

SUBJECT ID	HD	HD STD	BODY LEAN	BODY LEAN STD
1	1759.06	285.79	-46.058	141.73
2	1539.58	75.31	193.97	54.54
3	1866.12	15.96	52.41	20.19
4	1249.48	58.85	196.87	28.78
Test Avg	1367.581905	44.79	128.29	23.08

Table 5.11: Comparison of head depth and body lean features for test subjects before the exam was given. Test Avg is the average value recorded for that feature during all official test runs. Colors highlighted in gray area indicate a value larger that the test average.

Chapter 6

Conclusions

6.1 Conclusion

The combination of joint tracking, face tracking, and audio recording from the Kinect offers a comprehensive description of behavioral factors, allowing us to identify different postures, determine where the subject is looking, and calculate when the subject is talking. The reliability and level of detail the Kinect provides makes it well suited for continued automated behavioral analysis.

As expected, for visual attention tasks, eye gaze was the single strongest indicator of attention. Tracking where and when the subject is looking at the target is a straightforward calculation and direct measurement of attention.

The results from the analysis of body posture were also promising. Findings from this study suggest that an increased sample size and better random sampling of test subjects may improve the body posture regression analysis results. Even with the limitations of our sample set, we were able to find signs of correlation between posture and attention. While the single variable regression analysis showed inconsistent results for different features of posture, the multivariate regression showed a strong statistical significance between posture and attention. This indicated that one feature alone might not be enough to predict attention, but that considering multiple aspects of posture at the same time does provide a better degree of predictive power.

In summary, this research has shown that further investigation should be done in this field and that the methodology used for this study can be useful for future efforts

measuring attention using body posture or eye gaze direction.

6.2 Contributions

This research is the first to specifically address the prediction of user attention using the Microsoft Kinect. It demonstrated the first use of tracked joint data for measuring body posture and correlating it to user attention and is the first use of an objective measure of attention rather than a subjective measure of engagement for assessment of collected body posture features.

6.3 Limitations and Future Work

6.3.1 Limitations

CPTs and other sustained attention tasks are useful in measuring attention; however, the very nature of asking someone to participate in a task creates a situation where they are more likely to be attentive. This may not always be true for children or individuals with attention deficiencies, but this is the case for most non-attention disordered adults. While a CPT provides an objective measure of attention, its utility with normal adults is limited.

The Kinect system presented another limitation. The Kinect struggled to calculate head pose for several test subjects. This appears to be due to a configuration issue more than a limitation of the Kinect. Problems were only witnessed with taller test subjects who sat significantly higher than the test machine, resulting in large negative head pitch values. Placing the Kinect camera directly at the target (connected to the computer screen) instead of several feet above and behind the monitor, might have been enough to eliminate this problem.

6.3.2 Improvements

Results showed that the most important factor for predicting attention was identifying if the sensory organ for the target stimulus' modality was available and being

directed at the target. Using head pose for eye gaze estimation proved to be an effective method for identifying when an individual was looking at the desired target. A higher resolution decision could be made by enhancing the eye gaze estimation with a combination of face tracking and RGB image analysis, as proposed by [15], to track where the individual's eyes are looking.

Currently, the calculation for determining if a test subject is talking is very primitive. It considers at the volume level and reported source direction recorded by the Kinect microphone to decide if the subject is talking. This could be improved by using information from the Kinect face tracking API to determine if the subject's mouth is moving at the same time audio is being received.

6.3.3 Future Work

Future work should be conducted using the same test configuration with a more diverse group of test subjects. Focusing on young children or individuals diagnosed with ADHD would provide a larger spectrum of attentive behavior for regression analysis. The next step for this work is to build a system that runs in real time and to predict attention level based on the features discussed in this study. This work should also be extended to study predicting attention in an appropriate context, such as monitoring subjects watching a training video or participating in a Webinar.

While we have identified eye gaze as the best indicator of attention for visual tasks, this does not address the concept of covert attention, the act of focusing mental resources on an idea or thought. Being able to identify when one is covertly attending would help us predict scenarios such as 'daydreaming', where the subject is only appearing to be attentive. One possible approach to this would be to present the user with a stimulus that requires an "orienting response" that could be measured by the Kinect.

This research focused mainly on head and upper body positioning and movement, and did minimal work with tracking hand movement and position. There may be a great deal that can be determined from identifying this information and future work should include better tracking of hand movement and an analysis of its relationship to attention.

Appendix A

Attentional Tasks

A.1 Conners Continuous Performance Test

The Conners Continuous Performance Test (CCPT) is a (not-X) CPT, which asks the subject to respond every time a letter is presented that is not the letter X. The CCPT is 14 minutes long, during which time the subject is presented with a stimulus at 1, 2, and 4 second intervals. Each target is presented for 250 milliseconds. As of CCPT version 2 the values measured are[1]:

Omissions
Common CPT measurment recording the number of times the individual failed to respond to the target.

Commissions
Common CPT measurement recording the number of times the individual incorrectly responded when a non-target stimulus was presented.

Hit Reaction Time - Overall (Hit RT)
The average reaction time for correct responses across the entire test.

Standard Error - Overall (Hit RT Std Error)
Measures the consistency of response time for all responses (correct and incorrect). The larger this value is the more inconsistent the response times were.

[1]Measurements and descriptions derived from [3]

Variability of Standard Error

Similar to Hit RT Std Error, however instead of measuring the consistency of response time this value measures the consistency of the the subjects Std Error over 18 separate segments of the test.

Detectability (d')

A measure of the difference between the target and non-target distributions. This allows for assessment of the subjects ability to distinguish and detect X and non-X stimuli.

Response Style Indicator

Representation of an individuals response tendency. Some individuals tend to respond slower and less frequently in order to avoid commission errors while others tend to be less concerned with commission errors and more focused on responding to as many targets as possible.

Perseverations %

A response time less than 100 ms. This is often the result of mistakenly hitting the keyboard, responding late to the previous stimulus, or responding repeatedly without considering the stimuli.

Hit Reaction Time Block Change (Hit RT Block Change)

Measures a change in reaction time across the duration of the test. High values indicate a substantial slowing in reaction times while low values indicate the reaction time got faster during the test.

Standard Error by Block (Hit SE Block Change)

Measures changes in response consistency over the duration of the test. High values represent a loss of consistency while low values indicate an improved consistency.

Reaction Time by Inter-Stimulus Interval (Hit RT ISI Change)

Measures change in average reaction times across the different inter-stimulus intervals. An inter-stimulus interval is the time between the presentation of stimuli to the screen.

Standard Error by Inter-Stimulus Interval (Hit SE ISI Change)

Measures change in standard error across the different inter-stimulus intervals. An inter-stimulus interval is the time between the presentation of stimuli to the screen.

Further, these measures are broken into three categories summarizing what aspect they measure, Inattention, Impulsivity, and Vigilance. Inattention measures are Omission %, Comission %, Hit RT, Hit RT std. Error, Variability, Detectability (d'), Hit RT ISI Change, and Hit SE ISI Change. Measures of impulsivity are Comission %, Hit RT, and Perseverations %. The measures used for vigilance are Hit RT Block Change, and Hit SE Block Change.

A.2 Test of Variables of Attention

T.O.V.A Test of Variables of Attention is a specific continuous performance task. Like most CPTs it is generally used as part of a test battery to help diagnose children with ADHD. There are two versions of the T.O.V.A, one audio and one visual. This test is presented as a very simple computer game that runs for 22 minutes, during which time the subject is presented with two stimuli occurring at random intervals. Of the two stimuli, one is designated as a target and one is designated as a non-target. The subject is asked to respond to the target whenever it is presented on the screen by pressing a defined key on the keyboard. One advantage the T.O.V.A has over other continuous performance tasks is that it uses only two stimuli and those stimuli are geometric shapes so there is no concern over language barrier confusion. This test includes two sections, during the first part, known as the infrequent condition, targets randomly occur once for every 3.5 non-targets. This pattern is reversed for the second half of the task, or frequent condition, with targets appearing 3.5 times for every one non-target [14]. The following measures are recorded by T.O.V.A [2]:

Response Time

The time it takes for a person to respond to either a target or non-target measured in milliseconds. This is a measure of how quickly the person processes information and we typically find children with ADHD have a slower response time.

Response Time Variability

The consistency of response time, measured in milliseconds. This is an important measure for inattention because a change in response time could indicate a distraction or attentional resources directed elsewhere. [14] claim this is the most important measurement from the T.O.V.A. for identifying ADHD since

[2] Measurements and descriptions derived from [14]

subjects with ADHD tend to perform less consistently than average, sometimes answering faster and sometimes answering slower.

d' (d prime)
Measures how quickly a persons performance deteriorates over the course of the exam.

Errors of Commission This is the total number of times the subject responds incorrectly responds to the non-target when no response was desired. This is a common measurement in Continuous Performance Tasks which helps identify impulsivity in children.

Errors of Omission
This is the total number of times the subject did not respond when the target stimulus was presented. A common measurement in Continuous Performance Tasks which serves as a measure of inattention. The total number of omissions provides a measurement of how inattentive a subject was but taken independently each error of omission can be used as a marker indicating a point in time when the subject was not properly attending the screen.

Post Commission Response Time
Measures how the occurrence of a commission error affects the subjects response time immediately following the error. The expected result is that the subject will slow down and take more time in order to correctly identify the next target.

Multiple Response
The total number of times a person responded multiple times to a single target. The expected result is that this number will either be zero or very low.

Anticipatory Response
The total number of times a person responded so quickly to the target that it is likely they were guessing.

A.3 Psychomotor Vigilance Task

The Psychomotor Vigilance Task (PVT) is a sustained attention task that measures the reaction time of the subject in order to identify 'lapses'. A lapse is defined as a reaction time longer than what is expected to recognize the given target stimulus, typically around 500 milliseconds. PVT is frequently used as a measure of vigilance known to be sensitive to sleep loss [7]. In fact, NASA uses a version of this task for

their reaction self test to monitor the daily effects of fatigue on crew members on board the International Space Station (ISS) [16]. PVT results are generally interpreted as reflecting the arousal and attentional state of the individual [6].

PVT is a 5 or 10 minute task that presents a target stimulus to an individual with random inter-stimulus delay. Unlike the CPT the PVT is not concerned with omission and commission errors but rather it focuses purely on reaction time. The goal is to identify attentional lapses and use that information as a performance rating. Over the years the PVT has become acknowledge for its effectiveness at measuring performance effects due to fatigue or sleep deprivation. The following measurements are collected by PPVT:

Number of Lapses
: The total number of measured lapses during the test. This is the only official measure reported by the PVT and is used as a rating of the subjects performance.

Median Reaction Time
: Reported by some PVT implementations to provide more detailed statistical analysis of measured reaction times. This values represents the median reaction time for the exam.

Mean Reaction Time
: Reported by some PVT implementations to provide more detailed statistical analysis of measured reaction times. This values represents the average reaction time for the exam.

Reaction Time Standard Deviation
: Reported by some PVT implementations to provide more detailed statistical analysis of measured reaction times. This values shows how much variation there was in reaction time during the exam.

Appendix B

PCPT Report

```
--------------------------------------------------------
 Report for PEBL Continuous Performance Task (PCPT)
 Version 0.5
 http://pebl.sf.net
 (c) 2011 Shane T. Mueller, Ph.D.
 PEBL Version 0.12
 Sat Oct 20 12:50:38 2012
 Participant Code: 1
 Pause between blocks: 0
 Start Time: Sat Oct 20 12:36:37 2012
--------------------------------------------------------
Statistic        ISI:   1000 2000 4000 Pooled
--------------------------------------------------------
Correct Trials    114/120 111/120 110/120 335/360
Correct Targets   112/112 107/108 108/108 327/328
Correct Foils     2/8 4/12 2/12 8/32
Target Acc Rate   1 0.991 1 0.997
Foil Acc Rate     0.25 0.333 0.167 0.25
Commission Errors 6 8 10 24
Ommission Errors  0 1 0 1
Correct RT Mean   329.37 363.75 419.73 370.46
Correct RT SD     71.35 89.33 75.62 87.4
Error RT Mean     317.33 337.63 426.5 369.58
Error RT SD       41.49 42.19 67.4 72.73
Sensitivity (d')  0 -1.924 0 -2.068
Bias (beta)       0.804 0.764 0.858 0.806
--------------------------------------------------------
```

```
sub,block,trial,cond,targ,responded,corr,time,rt
1,1,1,2000,S,1,1,14299,392
1,1,2,2000,P,1,1,16299,264
1,1,3,2000,C,1,1,18299,305
1,1,4,2000,O,1,1,20299,376
....
1,4,75,4000,U,1,1,210305,426
```

Appendix C

PPVT Report

```
-------------------------------------------------------
 Report for PEBL Psychomotor Vigilance Task (PPVT)
 Version 0.3.  An Unprepared Serial Response Task (USRT).
 http://pebl.sf.net
 (c) 2008 Shane T. Mueller, Ph.D.
 PEBL Version 0.12
 Sat Sep 29 14:01:39 2012
 Participant Code: 1
-------------------------------------------------------
 Delay Count Median RT Mean RT SD RT
-------------------------------------------------------
1000 13 742 909.923 356.629
2000 16 1519.5 5484.38 15376.4
3000 4 1390.5 1317.25 561.523
4000 10 832 969.5 391.935
5000 18 804 1009.39 571.046
6000 14 881.5 2847.36 6920.1
7000 16 925 963.625 415.788
8000 15 619 882.867 510.153
9000 15 1004 1510.07 1717.56
-------------------------------------------------------
Too Fast:         0
Correct:          10
Lapse:            110
Sleep Attack:     1
-------------------------------------------------------
```

```
sub block trial ISI ISIbin abstime rt type
1 1 1 1269.72 1000 10209 901 3
1 1 2 5890.44 5000 14786 479 2
1 1 3 2700.74 2000 23562 693 3
1 1 4 6563.37 6000 29362 487 2
...
1 1 75 2395.86 2000 806225 2215 3
```

Appendix D

TOAV Report

```
--------------------------------------------------
 Report for PEBL Test of Attentional Vigilance
  (TOAV) Version 0.1
 PEBL Version 0.12
 Sat Oct 20 15:14:12 2012
 Participant Code: 2
 http://pebl.sf.net
--------------------------------------------------
Statistic           Half 1     Half 2     Pooled
--------------------------------------------------
Total Trials       320         320         640
Correct Targets    58          222         280
Correct Foils      239         54          293
Correct Trials     297         276         573
Commission Errors  9           18          27
Ommission Errors   14          26          40
Correct RT Mean    560         424         452
Error RT Mean      593         378         449
RT Mean            564         421         452
RT SD              189         131         157
Anticipations      0           1           1
Multiple Responses 1           2           3
--------------------------------------------------
```

```
sub trial targ toofast responded corr mult time rt
2   0     0    0       1         0    0    57329 563
2   1     1    0       1         1    0    59283 409
2   2     1    0       1         1    0    61238 454
2   3     0    0       0         1    0    63193 -1
...
2   74     0    0       0         1    0    202122 -1
```

Appendix E

Kinect Measures

E.1 Cross-sectional Features

`Distance Head` Proximity of head from display screen. Calculated as a distance (in meters) between the reported head joint and the location of the Kinect sensor.

`Distance Torso` Proximity of the torso from display screen. Calculated as a distance (in meters) between the reported Shoulder-Center joint and the location of the Kinect sensor.

`Shoulder Arrangement` Head tilt (left-right). Measured as the angle between the neck-head vector and the neck shoulder vector. This can also be calculated as a triangle consisting of the head, left shoulder, and right shoulder and then analyzed based on the lengths of the triangle's sides, angles between size, and total area of triangle.

`Gaze Direction` Heading (in degrees) of the subjects face relative to the location of Kinect sensor. Directly facing the sensor is treated a heading of zero degrees. This is calculated for both the x and y direction.

`Head Position` Position of the head in relation to the Kinect sensor. Measured in X,Y, and Z coordinate space with the Kinect sensor at 0, 0, 0.

`Hand Position` Position of each hand in relation to the Kinect sensor. Measured in X,Y, and Z coordinate space with the Kinect sensor at 0, 0, 0.

`Feet Position` Position of each foot in relation to the Kinect sensor. Measured in X,Y, and Z coordinate space with the Kinect sensor at 0, 0, 0.

`Volume of Motion` Total volume of motion. Total distance each joint travels, calculated as a summation of the distance traveled between each frame. For instantaneous values this is calculated as a moving window where each value recorded is the total amount of movement from the previous 10 seconds.

`Forward Body Lean` Angle and direction of back lean in relation to hip. This is the angle between the Y component of the Hip_Center-Shoulder_Center edge and the Kinect Y axis.

`Side Body Lean` Angle and direction of back lean in relation to hip. This is the angle between the X component of the Hip_Center-Shoulder_Center edge and the Kinect X axis.

`Slouch Factor` Degree of back curvature calculated from Hip_Center, Spine, Shoulder_Center, and Head joint positions.

`Out of Frame` Measures whether the individual is withing the sensors detection range. This can be caused by the subject leaving the designated area or by an unusual or complex posture that the Kinect is unable to detect.

`Talking` Measures whether the individual is talking during this particular moment in time. Talking is a good measure if inattention in situations where the subject is supposed to be attending the stimulus but not verbally responding to it. This generally indicates there is a distraction in present and that the subject is at least partially attending that over the target stimulus.

`# of People` Measures the number of people registered by the Kinect during a particular moment in time. The existence of other people alone might not be indicate a distraction but it increases the possibility. Combined with the the Talking metric this helps to measure the level of distraction surrounding the subject.

E.2 Aggregate Features

`Distance Head Mean` Average value of the Distance Head feature during the recorded time period.

`Distance Head Delta` Represents the variance in Head joint positions. Provides a measure of how far the head was moving during the designated time window.

`Distance Torso Mean` Average value of the Distance Torso feature during the recorded time period.

`Distance Torso Delta` Represents the variance in Spine joint positions. Provides a measure of how far the torso was moving during the designated time window.

`Average Arrangement` Average values for each metric reported in the arrangement triangle.

`Arrangement Delta` Represents the variance for each metric in the arrangement triangle. Allows for further analysis into which aspects of the head-should orientation were frequently changing.

`Gaze %` Percent of time gaze was directed at target.

`Movement Head` Total amount of movement by the Head vertex during the recorded time window. Calculated as a summation of the movement recorded for between each frame.

`Movement Hands` Volume of motion from hands. Total distance each hand travels, calculated as a summation of the distance traveled between each frame.

`Movement Feet` Volume of motion from feet. Total distance each foot travels, calculated as a summation of the distance traveled between each frame.

`Forward Body Lean Mean` Average value of the Forward Body Lean feature during the recorded time period.

`Forward Body Lean Delta` Represents the variance in Forward Body Lean angles.

`Side Body Lean Mean` Average value of the Side Body Lean feature during the recorded time period.

`Side Body Lean Delta` Represents the variance in Side Body Lean angles.

`Slouch Factor Mean` Average value of the Slouch Factor feature during the recorded time period.

`Slouch Factor Delta` Degree of back curvature calculated from hip, torso, neck, and head joint vectors.

`Out of Frame %` Percent of time the individual was considered out of the frame, either because the subject was not in front of the sensor or because the sensor could not detect the subject.

`Talking %` Percent of time the individual spent talking during the exam.

Appendix F

System Requirements

Hardware

- Microsoft Kinect for Windows
- Laptop 1 [Kinect] - Connected to Kinect for collection of posture and audio data.
- Laptop 2 [PEBL] - Run the PEBL sustained attention task exams. This will be the computer the test subject interacts with.
- Laptop 3 [PEBL] - Run only with Group 2. This is used to present visual distractions in the form of memory tasks the subject participate in.

System requirements Laptop 1 [Kinect]

Component	As Tested	Minimum Requirement
Processor	64 bit, 8 cores	dual-core, 2.66-GHz or faster processor
Hard disk	120 GB (internal SSD); 3TB (external USB 3.0)	2 TB for system drive. Hard disk space depends on how many recordings are taken. The drive written to must be capable of supporting system drive speeds. It is OK to use eSATA or USB 3.0 but USB 2.0 and below should be avoided.
Operating System	Windows 7	Windows 7 or Windows 8 - compatible graphics card that supports Microsoft DirectX 9.0 capabilities.

Table F.1: System requirements for Laptop 1 [Kinect]. *Note: Minimum requirements take from Microsoft Kinect SDK recommendations.*

Minimum required software

- Microsoft Kinect SDK 1.5
- .NET framework

System requirements Laptop 2 [PEBL]

Minimum required software

Component	As Tested	Minimum Requirement
Processor	32 bit, 1 core	32 bit, 1 core
RAM	2 GB	2 GB
Hard disk	60 GB	40 GB
Operating System	Windows XP	Windows XP or Windows 7

Table F.2: System requirements for Laptop 2 [PEBL].

- PEBL software

Appendix G

Multivariate Regression Results

G.0.1 PCPT

RT Features used as predictors for RT: AVGHD, HDGROUP, BODYLEAN, YAW, LOOK_AWAY_SUM, and PITCH.

Predictor	Coef	SE Coef	T	P
Constant	293.8	351.5	0.84	0.416
AVGHD	0.0836	0.2188	0.38	0.708
BODYLEAN	-0.0432	0.4239	-0.10	0.920
YAW	3.039	1.792	1.70	0.111
PITCH	1.731	2.056	0.84	0.413

S = 59.1871   R-Sq = 36.4%   R-Sq(adj) = 19.5%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	4	30116	7529	2.15	0.125
Residual Error	15	52547	3503		
Total	19	82663			

RT STD Features used as predictors for RT STD: AVGHD, HDGROUP, TALK

Predictor	Coef	SE Coef	T	P
Constant	91.30	54.37	1.68	0.121
TALK%	594.43	83.20	7.14	0.000
BODYLEAN	-0.9789	0.3498	-2.80	0.017
fleangroup	78.52	25.51	3.08	0.011
YAWD	-10.293	2.926	-3.52	0.005
LOOK_AWAY_SUM	0.031251	0.004522	6.91	0.000
PITCH	0.906	1.296	0.70	0.499
ROLLSD	17.062	4.892	3.49	0.005

S = 34.9136   R-Sq = 92.4%   R-Sq(adj) = 87.6%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	7	164016	23431	19.22	0.000
Residual Error	11	13409	1219		
Total	18	177425			

Further reduction of duplicate measures and poor contributors leaves us with

Predictor	Coef	SE Coef	T	P
Constant	-37.51	35.26	-1.06	0.304
TALK%	477.74	88.04	5.43	0.000
LOOK_AWAY_SUM	0.018755	0.003709	5.06	0.000
ROLLSD	19.426	5.977	3.25	0.005

```
S = 47.6061   R-Sq = 80.8%   R-Sq(adj) = 77.0%
```

```
Analysis of Variance
```

Source	DF	SS	MS	F	P
Regression	3	143430	47810	21.10	0.000
Residual Error	15	33995	2266		
Total	18	177425			

G.0.2 TOAV

RT Features used as predictors for RT: HDGROUP, LDROOP, LDROOP_SD, RDROOP_SD, YAWD, PITCHSD, ROLLSD

Removing duplicates and bad P-values.

Predictor	Coef	SE Coef	T	P
Constant	472.76	69.20	6.83	0.000
hdgroup	33.95	19.85	1.71	0.111
LDROOP	-2410	1083	-2.23	0.044
PITCHSD	3.177	5.132	0.62	0.547
ROLLSD	3.818	7.311	0.52	0.610

```
S = 44.7602   R-Sq = 63.1%   R-Sq(adj) = 51.8%
```

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	4	44616	11154	5.57	0.008
Residual Error	13	26045	2003		
Total	17	70662			

RT STD Features used as predictors for RT STD: HDSTDGROUP, SIDELEAN, SDIFFSD, LDROOP, LDROOP_SD, RDROOP_SD, ROLLSD

Predictor	Coef	SE Coef	T	P
Constant	23.32	21.31	1.09	0.294
hdstdgroup	-27.88	15.88	-1.76	0.103
TALK%	112.93	58.73	1.92	0.077
RDROOP_SD	5492	1620	3.39	0.005
ROLLSD	4.222	3.451	1.22	0.243

S = 36.8008   R-Sq = 60.5%   R-Sq(adj) = 48.3%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	4	26934	6734	4.97	0.012
Residual Error	13	17606	1354		
Total	17	44540			

G.0.3 PPVT

RT Features used as predictors for RT: BODYLEAN, SDIFF, YAWD, RDROOP, PITCH, ROLLSD

Predictor	Coef	SE Coef	T	P
Constant	209.48	61.07	3.43	0.005
BODYLEAN	0.5405	0.3926	1.38	0.194
SDIFF	0.6004	0.5659	1.06	0.310
YAWD	-4.836	6.446	-0.75	0.468
RDROOP	667.2	494.6	1.35	0.202
PITCH	1.928	1.515	1.27	0.227

ROLLSD	11.479	5.267	2.18	0.050

S = 35.4602   R-Sq = 48.6%   R-Sq(adj) = 22.9%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	6	14270	2378	1.89	0.164
Residual Error	12	15089	1257		
Total	18	29359			

RT STD Features used as predictors for RT STD: HDGROUP, TALK%, BODYLEAN, LDROOP,

Predictor	Coef	SE Coef	T	P
Constant	136.53	31.79	4.29	0.001
hdgroup	-15.74	15.67	-1.00	0.331
TALK%	171.54	77.47	2.21	0.043
BODYLEAN	-0.1570	0.1744	-0.90	0.382
LDROOP	-660.1	536.1	-1.23	0.237

S = 34.0259   R-Sq = 42.4%   R-Sq(adj) = 27.0%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	4	12771	3193	2.76	0.067
Residual Error	15	17366	1158		
Total	19	30137			

G.0.4 TOAV+PPVT

RT Features used as predictors for RT: AVGHD, HDGROUP, BODYLEAN, FLEANGROUP, SDIFF, YAWD, PITCH, PITCHSD, ROLLSD

Predictor	Coef	SE Coef	T	P
Constant	47.6	268.6	0.18	0.860

AVGHD	0.1899	0.1768	1.07	0.291
fleangroup	18.72	21.91	0.85	0.399
YAWD	4.420	6.164	0.72	0.479
PITCH	0.969	1.541	0.63	0.534
ROLLSD	5.763	5.602	1.03	0.312

S = 62.5725   R-Sq = 27.6%   R-Sq(adj) = 15.9%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	5	46227	9245	2.36	0.063
Residual Error	31	121375	3915		
Total	36	167602			

RT STD Features used as predictors for RT STD: TALK%, LDROOP, LDROOP_SD, RDROOP_SD, ROLLSD

Predictor	Coef	SE Coef	T	P
Constant	119.86	27.25	4.40	0.000
TALK%	116.02	52.45	2.21	0.034
LDROOP	-1028.4	401.7	-2.56	0.015
ROLLSD	3.582	2.272	1.58	0.125

S = 41.1148   R-Sq = 28.0%   R-Sq(adj) = 21.4%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	3	21687	7229	4.28	0.012
Residual Error	33	55784	1690		
Total	36	77471			

RT STD 2 Features used as predictors for RT STD: TALK%, LDROOP, LDROOP_SD, RDROOP_SD, ROLLSD

Predictor	Coef	SE Coef	T	P
Constant	127.46	33.81	3.77	0.001
SDHD	0.8184	0.3051	2.68	0.013
AVGVOL	-3101	1035	-2.99	0.006
TALK%	504.7	132.4	3.81	0.001
BODYLEAN	-0.2084	0.1562	-1.33	0.194
BODYLEANSD	-3.6087	0.9792	-3.69	0.001
SIDELEANSD	5371	1496	3.59	0.001
SDIFFSD	-0.5704	0.5560	-1.03	0.314
YAWD	-8.337	3.569	-2.34	0.027
ROLLSD	7.955	3.545	2.24	0.034

S = 33.8274   R-Sq = 61.6%   R-Sq(adj) = 48.3%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	9	47691	5299	4.63	0.001
Residual Error	26	29752	1144		
Total	35	77443			

Predictor	Coef	SE Coef	T	P
Constant	85.67	18.17	4.72	0.000
SDHD	0.6768	0.2866	2.36	0.025
AVGVOL	-2862	1019	-2.81	0.009
TALK%	493.8	131.4	3.76	0.001
BODYLEANSD	-2.9684	0.9083	-3.27	0.003
SIDELEANSD	4383	1378	3.18	0.004
YAWD	-7.452	3.533	-2.11	0.044
ROLLSD	8.359	3.485	2.40	0.023

S = 34.2345   R-Sq = 57.6%   R-Sq(adj) = 47.0%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	7	44627	6375	5.44	0.001

```
Residual Error  28  32816  1172
Total           35  77443
```

Bibliography

[1] T.S. Bearden, J.E. Cassisi, and J.N. White. Electrophysiological correlates of vigilance during a continuous performance test in healthy adults. *Applied psychophysiology and biofeedback*, 29(3):175–188, 2004.

[2] J. Bidwell and H. Fuchs. Classroom analytics: Measuring student engagement with automated gaze tracking.

[3] C.K. Conners and MHS Staff. Conners' continuous performance test ii (cpt ii v. 5). *Multi-Health Systems Inc., North Tonawanda, NY*, 2000.

[4] Microsoft Corporation. *Human Interface Guidelines v1.5.0*, 2012. Accessed: 10/14/2012.

[5] S. D'Mello and A. Graesser. Automatic detection of learner's affect from gross body language. *Applied Artificial Intelligence*, 23(2):123–150, 2009.

[6] S.P.A. Drummond, A. Bischoff-Grethe, D.F. Dinges, L. Ayalon, S.C. Mednick, MJ Meloy, et al. The neural basis of the psychomotor vigilance task. *SLEEP-NEW YORK THEN WESTCHESTER-*, 28(9):1059, 2005.

[7] P.L. Franzen, G.J. Siegle, and D.J. Buysse. Relationships between affect, vigilance, and sleepiness following sleep deprivation. *Journal of sleep research*, 17(1):34–41, 2008.

[8] J. Geigel. Teaching animation in second life. In *SIGGRAPH 2009: Talks*, page 64. ACM, 2009.

[9] W.S. Geisler and L. Cormack. Models of overt attention. *Oxford handbook of eye movements*, 2011.

[10] J.F. Grafsgaard, K.E. Boyer, E.N. Wiebe, and J.C. Lester. Analyzing posture and affect in task-oriented tutoring. In *Proceedings of the 25th Florida Artificial Intelligence Research Society Conference*, pages 438–443, 2012.

[11] L.M. Greenberg and I.D. Waldmant. Developmental normative data on the test of variables of attention (tova). *Journal of Child Psychology and Psychiatry*, 34(6):1019–1030, 2006.

[12] W.F. Horn, C.K. Conners, K.C. Wells, and D. Shaw. Use of the abikoff classroom observation coding system on a children's inpatient psychiatric unit. *Journal of psychopathology and behavioral assessment*, 8(1):9–23, 1986.

[13] W. James, F. Burkhardt, and I.K. Skrupskelis. *The principles of psychology*, volume 1. Harvard Univ Pr, 1981.

[14] MD Lawrence M. Greenberg. Attention deficit disorders and the test of variables of attention. `http://www.biof.com/onlinestore/tova/attention.asp`. Accessed: 10/14/2012.

[15] Kenneth Alberto Funes Mora and Jean-Marc Odobez. Gaze estimation from multimodal kinect data.

[16] NASA. Reaction self test. `http://www.nasa.gov/mission_pages/station/research/experiments/Reaction_Self_Test.html`. Accessed: 30/09/2012.

[17] S.L. Nichols and D.A. Waschbusch. A review of the validity of laboratory cognitive tasks used to assess symptoms of adhd. *Child Psychiatry & Human Development*, 34(4):297–315, 2004.

[18] O. Patsadu, C. Nukoolkit, and B. Watanapa. Human gesture recognition using kinect camera. In *Computer Science and Software Engineering (JCSSE), 2012 International Joint Conference on*, pages 28–32. IEEE, 2012.

[19] J. Sanghvi, G. Castellano, I. Leite, A. Pereira, P.W. McOwan, and A. Paiva. Automatic analysis of affective postures and body motion to detect engagement with a game companion. In *Proceedings of the 6th international conference on Human-robot interaction*, pages 305–312. ACM, 2011.

[20] Rainer Stiefelhagen. Tracking focus of attention in meetings. In *Multimodal Interfaces, 2002. Proceedings. Fourth IEEE International Conference on*, pages 273–280. IEEE, 2002.

[21] O. Tucha, S. Walitza, L. Mecklinger, T.A. Sontag, S. Kübber, M. Linder, and KW Lange. Attentional functioning in children with adhd–predominantly

hyperactive-impulsive type and children with adhd–combined type. *Journal of neural transmission*, 113(12):1943–1953, 2006.

[22] P. Zimmermann and B. Fimm. A test battery for attentional performance. *Applied neuropsychology of attention. Theory, diagnosis and rehabilitation*, pages 110–151, 2002.