Improving Task Performance in an Affect-mediated Computing System

Vivek Pai

April 30, 2012

Abstract

Computing systems have traditionally relied on purely quantitative means of assessing users. Software may record facts about users such as their click behaviors, their accuracies, and their rates of recurrence, and may then use this acquired data to make predictions about their behaviors and preferences. However, users' inputs into a machine, sequences of key presses and mouse movements, do not capture all of their desires and needs. The machine may never know that a particular user enjoys the glow of a button and its placement among others elements of an interface. Furthermore, it cannot perceive a user's discomfort during accomplishing a task; it may simply translate the user's high accuracy in hitting various steps of a process into a mastery of the task. The machine, which might be intelligent, may not sense the user's uncertainty of his or her actions, expressed by widened eyes, a stutter in voice, and a hand clasped behind the neck.

In this project, we explore how an affect-mediated system, a computing system that adapts its actions and behavior to the emotional state of its users, can improve their abilities to complete tasks and meet their goals. In particular, we apply facial expression recognition, one method for estimating a human's emotional state, to a children's game, giving it the ability to adjust its difficulty based on its player's perceived unease. Through experimentation with this game, we determine whether affect-mediation helps users achieve their goal of winning the game, and whether in general affect-mediated systems can aid in user task completion.

Our study consisted of two conditions: affect (the game would adjust its difficulty based on the player's emotion) and control. Participants in the affect condition tended to have a higher catch/total ratio (the number of items caught by the participant divided by the total number of items spawned in the game). They also had less struggles (a catch followed by one or several misses) and these struggles were more spread out than in the control condition, suggesting a higher level of engagement.

Acknowledgements

This thesis would not be possible without the support of many people. The author wishes to express his gratitude for his mentor Dr. Raja Sooriamurthi (Associate Teaching Professor, Information Systems Program, Carnegie Mellon University) who provided much assistance and advising throughout the process. He would also like to thank Dr. Yaser Sheikh (Assistant Research Professor, Robotics Institute, Carnegie Mellon University) for pointing him in the right direction in the research aspect of the project and for suggesting the work of Jason Saragih (Research Scientist, CSIRO ICT Center Computer Vision laboratory). Saragih's FaceTracker was an indispensable component of the core system the author designed and developed and the author greatly appreciates his work.

The author would like to thank Dr. Sharon Carver (Director Children's School, Carnegie Mellon University) for allowing him the opportunity to conduct his study at the Children's School. He also wishes to convey thanks to Ms. Allison Drash (Administrative Coordinator Children's School, Carnegie Mellon University) for helping schedule all of his various testing appointments at the School, and the Kindergarten and Pre-school staffs for facilitating the author in his interaction with their children.

The author is also grateful for the Information Systems Department at Carnegie Mellon University for helping him in all of his endeavors.

Lastly, the author wishes to express his gratitude to his family for their understanding & support throughout his four years at Carnegie Mellon University.

Contents

Part I

Introduction

THUD. A grunt. Traces of coffee run down the cup. WHACK. A dark puddle envelopes the base. Again a grunt, and then a moan. His upper lip rises, eyebrows collapse in, breathing exaggerates. Body presses forward, bottom on seat edge, hand clasping mouse. SLAM. Silence.

Fingers loosen hold. Back straightens. A sigh and then a deep breath. Puddle's quivers subside, coffee drips to a halt.

Computer: "Congratulations, you have successfully placed an order for 'nylon mittens'. Have a wonderful day."

The computer records the details of the user's transaction and updates the user's profile. It readjusts its internal model of the user's item preferences, favorite categories, and time of day purchases in order to predict the user's next purchase. The machine moves closer with each transaction to developing a thorough understanding of its user's behavior.

The user visits the order site again the next week. The computer, having adapted its model of the user, places the user's favorite categories in a box towards the top of the page. The user clicks on a category from the box, browses through various subsections, and then finds an item to purchase. He proceeds to the order page where he must fill in his billing and shipping information. Right before he hits the submit button, he glances at his shopping cart. He realizes that he had accidentally selected an accessory and not the actual product he wanted.

THUD. WHACK. SLAM

Instead of replacing the wrong item from his shopping cart with the correct one, the user abruptly closes the order window and storms off from his workspace. The computer tracks the incomplete order purchase and readjusts its model of the user. It decides to display the favorite categories box for the user the next time he logs back in and chooses to also emphasize in this box the category that the user had clicked.

But this adjustment may not exactly help the user whose impaired spatial reasoning hinders his capacity to relate items such as buttons on a webpage. The user consis-

tently struggles to determine which "Add to Cart" button to use to add the product to the cart. In most order sessions, he mistakenly adds the wrong one and must go back through the order process to replace it with the right product. Meanwhile, the machine takes each modification to the shopping cart as simply the user changing his mind. It may or may not respond to the user's wavering decisions at the time, but ultimately it tracks various factors of the user in order to improve his ordering efficiency. Unfortunately, the true issue lies in the user's ability to choose the right button and not in his speed in entering into a commonly accessed category.

Thus, it would seem that two realities coexist: what the computer believes it understands about the user and what the user truly thinks and experiences. The computer waits patiently for each mouse click and keystroke up until the moment the user confirms the order. But more than just clicks and keystrokes have transpired from the user's perspective. The user perceives and processes the various images and text that appear on screen. He concocts a set of actions to take next, chooses one to perform and then waits for a response from the computer, a new visual or auditory cue which will then restart his cognitive cycle. Additionally, he may spend more time at a particular step of the cycle. For example, he may bathe in a bed of thought, weighing each alternative choice to decide the better fit. Ultimately, the user participates in more than just those events transmitted directly into a computer, and thus the computer may construct an incomplete assessment of the user.

The Lagging Machine

The machine misses many details about the user, simply treating him as a static, predictable entity. The user's state is one of many associated with hardwired rules embedded in code. If his state confirms that he has passed all three steps of a process, then he may move to the final step. But this state says nothing about the user's confidence in the process, for humans are far more dynamic. Our limits of cognition surpass those of other animals. We do not just move around in our spaces, foraging for food, searching for means for survival. We bend our surroundings, twist it into art forms. We breathe

it in and react to it in many ways. Our state can change in an instant with every new event or ordeal. We may even be in multiple states in some time period and in some cases those states may vacillate back and forth. An assessment of a user through solely those moments he taps a key may be an oversimplification of the user's behavior.

Further attempts to model the dynamic nature of human cognition are still not sufficient to understand users. Artificial intelligence affords a machine the ability to learn and construct new rules to map factors from the environment to decisions. In this sense, the machine's actions based on a user's state may change with new information that will subsequently modify an existing rule set. But these rules still capture only the results of keystrokes and mouse clicks – they do not express every instance of a human's behavior expressed in-between or simultaneously with his input into the machine. More so, a rule may "fit" the user in his initially captured state, but the user's wavering mind may deviate from this initial state, leaving the rule inconsistent and stale. Even if in that particular instance the artificially intelligent agent made a correct assumption of the user's behavior, the user may simply proceed in a completely different direction than the agent had anticipated, and the machine will lag behind the user's new state.

In particular, what prompts the variability of these states is the user's sentiments, his feelings and moods in situations. How a user feels about his choices may sway him towards one or another. Logically, if a person wanted to submit a report by some deadline, he would print out his document and turn it in. But fatigue might intervene and he may feel too exhausted to walk over to deliver his report, and so, might abandon the idea altogether. Furthermore, his emotional state may further manipulate his actions after a choice has been made. A person's immense joy may blind him from reading through all the guidelines of an instruction manual, and thus he may make mistakes in his usage of the product. In this case, feelings allow for human infallibility, the ability to tread irrationally and unpredictably into one of many states of err. There can be infinitely many different sequences of states the user can be in at any given time period, for every thought generates a feeling, which may subsequently motivate the mind and propel it into action.

Thus, we might decide to simply tailor the artificially intelligent program's actions to predictions of how the user's state will change over time. The program would generate as many possible sequences of the user's state, going beyond those which the artificial agent can logically predict, and then just choose one sequence to inform its action. But then, the question now becomes, how do we make this choice? We can't just decide arbitrarily which sequence to assume or else our machine would not be intelligent, but rather, random. We also cannot process each and every possible sequence of states, for to capture as many humanly possible infinite sequences, we have to generate close to as many sequences. A machine cannot judge every sequence as that would be computationally prohibitive on top of the already daunting task of generating these intermediate state sequences. Thus, in order to have a program with such foresight so as to predict the sequence of states a user will traverse during their interaction, the program must generate only a smaller set of possible predictions; consequently, this set may not express all of its user's behavior or personality. Attempting to predict the user's changing behavior might not work well if we have to guess from an infinite set of possibilities.

Therefore, we may find relief in the user informing the program of his internal state instead of the program attempting to predict it. We cannot exactly ask the user questions through the process for there may be infinitely many possible questions (if we wish to grasp a hold onto any arbitrary thought that flows through the user's mind). Furthermore, some of these questions simply might not apply to one's thoughts. For example, it might not be appropriate to question the user on his level of fear if a sequence of blissful events run through his mind. Requiring the user to update the system on the state of his mind might not be practical either as it would simply disrupt the user's task flow. Thus, we need to be able to dissect users with minimal intervention in order to fill the gap in the program's knowledge of the user.

Though, we might not need any intervention in the first place as a user's behavior is already visible through some external representation. His subtle quivers and bouts of perspiration make up the physical realization of his internal state, his mind's current emotional frame. As this frame advances, the rest of his body signature shifts to indi-

cate this change. In particular, we can simply observe the user's emotional expression, monitor these slight changes in his presentation and vocal tone.

And these observations reliably hint at one's inner state. According to research by Paul Ekman in the late 1960s and 1970s, all humans express the same seven basic emotions consistently (Ekman, Friesen, & Sorenson, 1969; Ekman & Friesen, 1969). Joy usually compels one to smile and raises the pitch of one's voice, while a frown (and in some cases tears) accompanies a feeling of sorrow. Moreover, cognition drives these emotions. A user who might perceive a situation as impossible may begin to feel anxious and his jaw might drop, his heart might beat faster, and his arms will shake in response. The feelings map back to a person's inner thoughts as they are constructed by them. Additionally, these feelings may simply control his thoughts, for the fear of the impossible might hinder any attempts by the user to act beyond his limits. Thus, the analysis of one's emotional state may provide an estimation of his cognitive complex.

In order to intelligently guess at a user's emotional state, we must observe the fragments that constitute his emotional makeup. Machines already have eyes and ears through built-in (or attached) cameras and microphones. Through computer vision and speech analysis techniques, they can leverage these virtual senses to pick up the variations in their users' physique and discourse. Moreover, they may hypothesize the user's emotional state directly. Ekman in his research on the seven basic emotions discovered that each emotion maps to a specific facial expression and vocalization and that these pairings were universal (Ekman et al., 1969; Ekman & Friesen, 1969). Thus, algorithms can convert just sight and sound into reasonable estimates of the user's emotional state.

We may further use these algorithms to complement our program's assessment of its user's state. We simply turn on the machine's virtual senses and wait for every expressional change in the user. Even after the program has decided upon an action, the virtual senses can fine tune this action and maybe even avert it in order to adapt to the user's new internal state. The program stays in sync with the user's state changes,

responding immediately to those needing care and attention. Once a passive entity, the program now with its own pair of eyes and ears engages the user in their interaction.

Thus, we see in a theoretical lens that we can augment a machine with affect detection to better support users. The rest of this thesis will expand on this idea by examining the role of affect in our execution of tasks, how specifically we can detect these various affective states, and how this detection can improve a user's task performance. We begin with a brief overview of the concept known as emotion, starting with the biological underpinnings of cognition and emotion, then expanding on Ekman's own discoveries regarding the basic emotions, and ending with the interplay between emotion and cognition. We then introduce facial expression recognition, one popular method of emotional detection. Lastly, we look into the burgeoning field of affective computing, "computing that relates to, arises from, or influences emotions" (Picard, 1997). Here we assess the effectiveness of affective programs through my own affective system, a children's computer game that adjusts its difficulty based on the player's emotional state. In our study using this affect-mediated game system, **we expect that participants will have better performance playing the game when it responds to their emotions than when it completely ignores them**. Through this broad view of the concept of affect-sensitive machines, we hope to convey that these types of computing systems are very feasible given modern technology and may advance us into the next generation of interactive systems.

Part II

What is Emotion?

1 Biology

Neurons

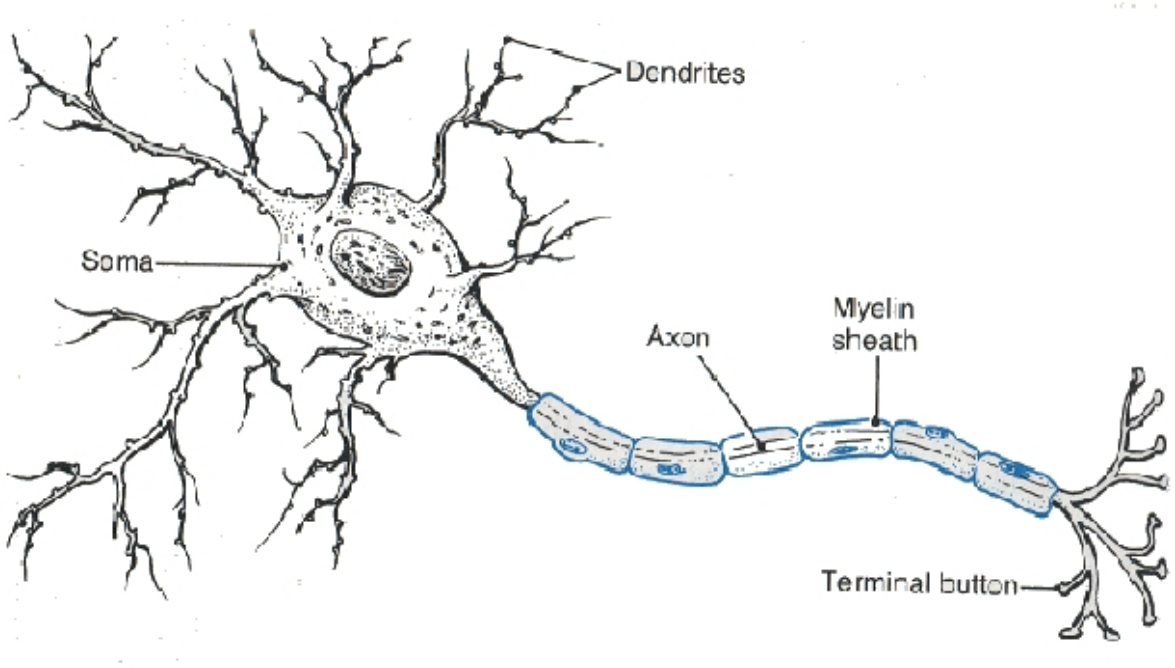


Figure 1: A neuron
http://www2.cedarcrest.edu/academic/bio/hale/bioT_EID/lectures/tetanus-neuron.html

Neurons compose cognition at its lowest level. They each consist of a body (the soma), an axon (a long structure that extends from the soma), and several dendrites (tinier structures surrounding the soma) (Figure 1). Neurons communicate with each other through chemical signals known as neurotransmitters. Various ion channels line the membranes of neurons, through which ions such as potassium, sodium, chloride, and calcium pass in and out of the cell, creating an ion concentration difference across the membrane. The difference results in a cross-membrane voltage. When this voltage changes (due to a difference in concentration), an event known as an action potential ensues, causing the voltage to immediately rise and then drop, resulting in the opening and closing of ion channels and the subsequent release of neurotransmitters at their synapses (Figure 2). Neurotransmitters jump from axon to dendrites through these synapses and may further excite receiving neurons, compelling them to send further signals to adjacent neurons. Neurons take inputs, process those inputs, and

fire outputs to the next neuron, simulating a message delivery system that extends throughout the entire body.

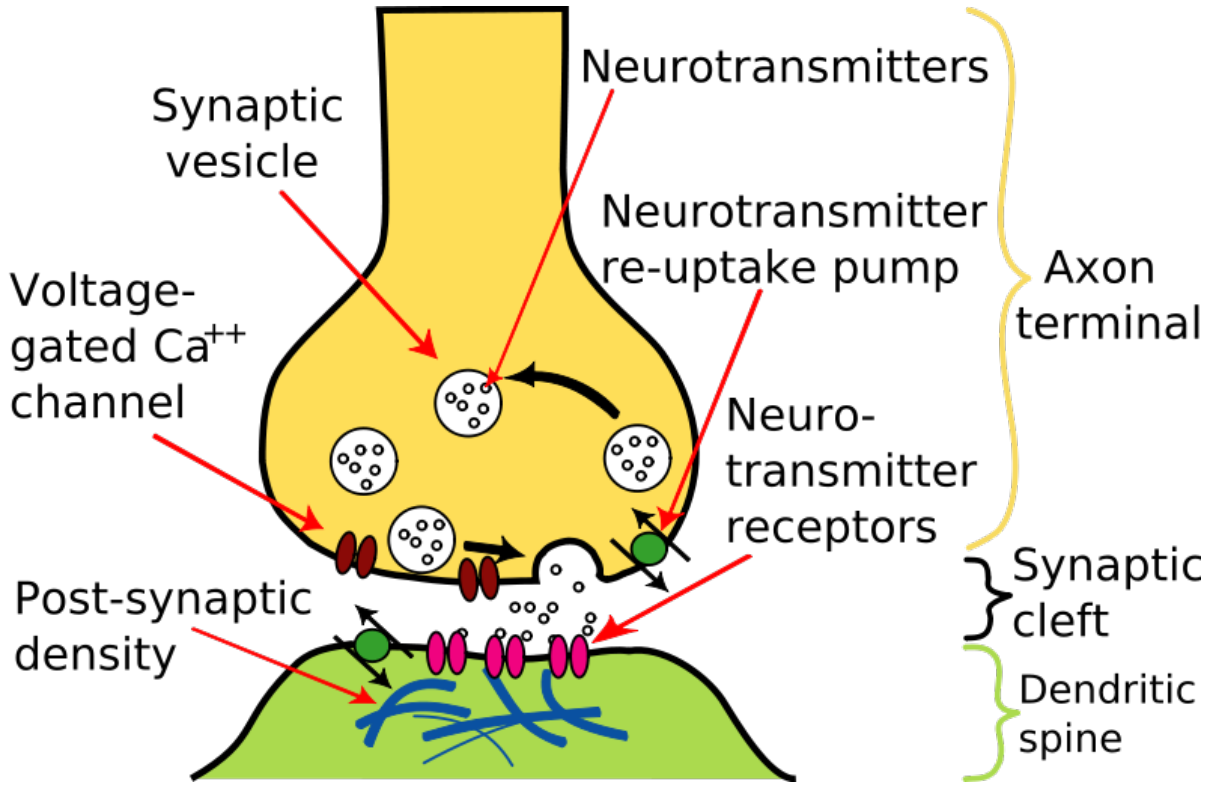


Figure 2: The synapse of a neuron.
http://cephalove.blogspot.com/2010_05_01_archive.html

Neurons can be classified into three types: sensory, motor, and interneurons. Sensory neurons respond to touch, sound, light, and other stimuli, and then send signals to the brain. Motor neurons, on the other hand, receive signals from the brain to generate muscle contractions and command adjacent glands. Interneurons simply bridge neurons. All three types play a role in conducting the various cognitive processes that occur in a human.

Key Features of the Brain

The brain comprises billions of neurons that signal and communicate with each other in order to carry out distinct cognitive processes. It is generally decomposed into three main parts: the cerebrum (the largest part, located towards the front), the brainstem (located in the middle of the brain), and the cerebellum (located at the back). Fur-

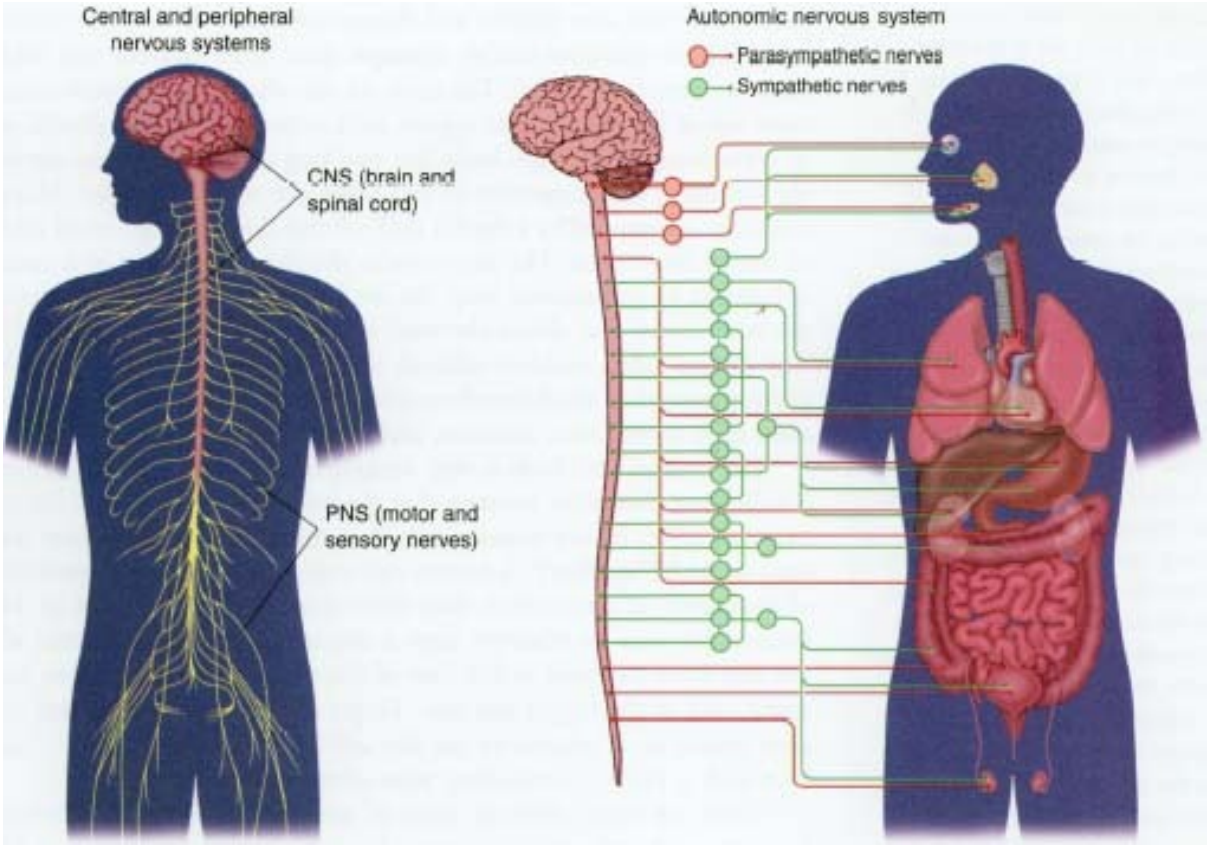


Figure 3: The nervous system comprised of both the Central Nervous System (CNS) and the Peripheral Nervous System (PNS).
http://www.humanillnesses.com/Behavioral-Health-A-Br/The-Brain-and-Nervous-System.html#b

thermore, it is part of the central nervous system (CNS) and communicates with the peripheral nervous system (PNS), which encompasses neurons in the rest of the body (Figure 3). The brain sends signals to the motor neurons in the PNS and receives signals from PNS sensory neurons. The analysis and the response are determined by the interplay of various structures within the brain itself.

The cerebellum commands of our very basic functions, such as our ability to balance and coordinate. It stores skills we don't consciously think about. For example, it allows an experienced piano player to automatically hit the right keys without looking at his hands and without thinking of the key-note mapping. The cerebellum takes care of the explicit actions for us during the task, affording us the ability to multitask (to hold a conversation while walking, for example).

The cerebellum is connected to the brain stem, another region of the brain that regulates our basic functions. It directs eye and voluntary body movement, and it is involved in maintaining vital functions such as breathing and our heart rate. See (Table 1) for a description of sub structures and their functions.

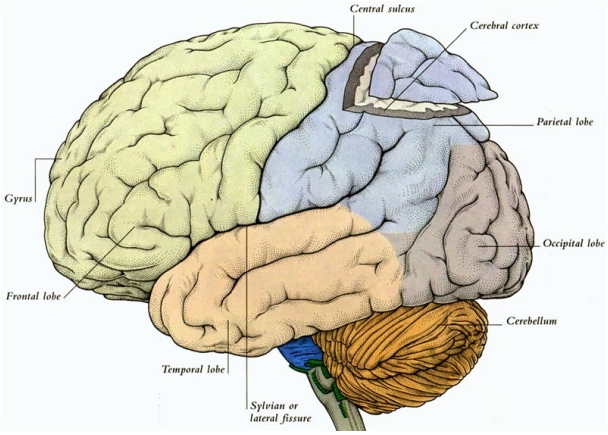


Figure 4: The brain.
http://tlhung.blogspot.com/2010/03/part-4-improving-human-memory.html

Table 1: Brain stem sub structures

Structure	Function
Pons	sensory analysis, regulates consciousness and sleep, coordinates eye movements/balance
Medulla	maintains vital functions like breathing, heart rate, swallowing
Spinal cord	maintains vital functions like breathing, connects CNS to nerves in rest of body

While the brain stem and the cerebellum are responsible for our more automatic cognitive processes, the cerebrum provides us with higher cognition capabilities. It lets us think and feel. It moves us beyond the moment, allowing us to recall events from the past and to hypothesize the future. It also gives us choices. In a room, you perceive six differently colored doors and you must choose one door to walk through. You remember your negative experience the last time you chose the blue door, so you know you must not choose the blue door this time, and consequently, you decide to walk through another door instead. Much of our higher level cognitive processing occurs in four different quadrants of the cerebrum: the frontal lobe, the parietal lobe, the occipital lobe, and the temporal lobe (these are described much more in detail in Table 2) (Figure 4). Furthermore, within the cerebrum are a set of structures that form

the limbic system (see Table 3). The limbic system influences both the endocrine system (system of glands that secrete hormones) and the autonomic nervous system (part of the PNS, influencing several non-voluntary processes). For the most part, the cerebrum provides us consciousness and the ability to deviate from a natural course of life.

Table 2: Lobes of the cerebrum

Structure	Function
Frontal lobe	reasoning, planning, problem solving, voluntary movement, concentration, judgment of stimuli
Parietal lobe	spatial/visual perception, perception and discrimination of sensory stimuli (from vision and hearing), processing the sense of touch, pain, and temperature
Occipital lobe	visual processing and interpretation
Temporal lobe	perception and recognition of auditory stimuli, information retrieval from memory, speech, sequencing/organization

Emotion in the Brain

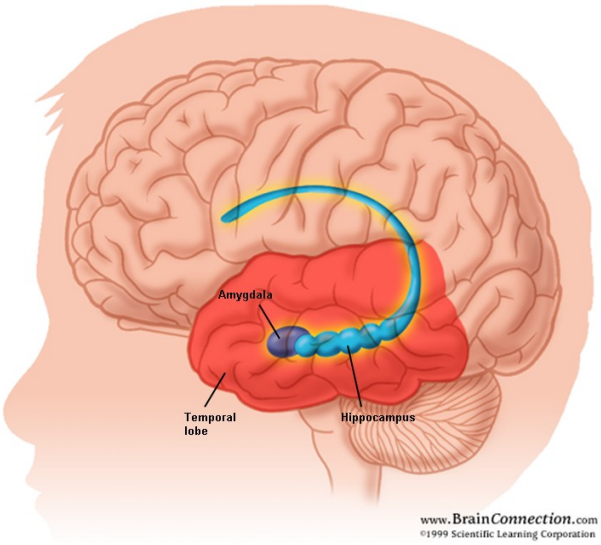


Figure 5: The amygdala.
http://www.positscience.com/human-brain/image-gallery/brain-anatomy-images?page=1

In particular, the amygdala (Figure 5) plays a major role in processing emotional reactions. It receives input from various sensory systems and sends impulses to the hypothalamus to release acetylcholine neurotransmitters (chemical signals), which activate the sympathetic nervous system (SNS). Part of the autonomic nervous system, SNS initiates the body's flight or fight response, dilating the pupils and bronchioles and in-

Table 3: Components of the Limbic System

Structure	Function
Thalamus	processes most sensory information (except olfactory) before sending it to the cerebral cortex, pain sensation, alertness, attention, memory
Hypothalamus	controls autonomic nervous system, governs emotional behavior, thirst, homeostasis, controls pituitary gland and influences endocrine system
Hippocampus	important for learning, converts short term (working) memory into long term memory, recalls spatial relationships
Temporal lobe	perception and recognition of auditory stimuli, information retrieval from memory, speech, sequencing/organization
Amygdala	emotion, plays role in regulating memory consolidation

creasing the rate and force of heart contraction (the parasympathetic nervous system is the dual to the SNS, relaxing and putting the body to rest). The amygdala also sends impulses to the facial nerves, causing facial muscles to contract and to thereby generate various facial expressions (similarly, facial expressions as stimuli will activate the amygdala). Furthermore, it may influence the nucleus accumbens, another part of the limbic system that can process both incentives and pleasant stimuli and can influence addictive behaviors (Winkielman & Trujillo, 2007, p. 85). Lastly, the amygdala plays a great role in regulating memory consolidation (the conversion of short term memory to longer lasting memory); moreover, it forms and stores memory for stimuli and events that are emotionally arousing. Thus, the amygdala can be considered the "heart" of the brain for its great influence on other structures in emotional response.

Emotional Expression

The actual emotional state of a human cannot be determined; rather, it can be guessed or approximated (with high accuracy). The face is one medium through which one can determine another's emotional state. However, the amygdala in its processing of emotion, influences structures that send signals throughout the body which contract muscles and stimulate sweat glands. According to Rosalind Picard, the director of the Affective Computing Research Group at the MIT Media Lab , variance in finger pressure and foot pressure may express a wide range of emotional states, and one's heart rate may assist in differentiating between two different states (Picard, 1997, p. 5). Thus, regions all over the body contribute to the expression of an emotional state.

Furthermore, this expression may appear independently of other processes. Picard states that the "will and the emotions control separate paths" (Picard, 1997, p. 5). We might consciously tell our body to move in a particular fashion, but our emotionally aroused brain may concurrently signal our motor system to behave differently, thereby influencing the actual movement. She further notes an observation from neurological literature of a patient paralyzed on one side: telling the patient to smile will result in the patient's smile appearing on only his active side while cracking a funny joke will result in a full smile (Picard, 1997, p. 5). Thus, the actual smile appears through the genuine feeling of joy, suggesting that affective states may sidestep any cognitive limitations. The emotional state and a conscious coexist, potentially interacting with each other.

2 Affect and Cognition

Instances of affect and cognition manifest independently, but may also steer each other. Cognitive processes can motivate affective responses. One understands the consequences of being late to work, and in the same moment, the person's face may warp into fear, his skin may drench with sweat, and his heart might pound faster. One's vocal tone might rise and his face might glow upon receiving a high grade for an as-

signment. Upon understanding a situation, the brain may send signals all over the body to enact a particular emotion.

On the other hand, a person's affective state may drive these cognitive processes. They can determine what the brain will attend to when processing sensory information and how this information will be processed, judged, and remembered. They can also influence both the decision to carry out a task and the actual performance on the task.

Attention grabbers

Stimuli surround our physical world. In the basic form, there are visual details that vary in color and intensity, auditory details varying in tone, pitch, and duration, tactile details (in coarseness and temperature), olfactory details (in freshness, purity), and taste details (in sweetness and saltiness). The brain integrates the basic sensory information into more complex forms. Pink birds chirp in the distance. Steam emits from a moist chicken, sliding off the lavender plate. There are many possible stimuli at any given moment that the brain must make a choice to attend to (if it should attend to any). One stimuli will win for attention either because it is emphasized in the environment (a blinking light) or because the brain places greater priority on this object based on its relevance to particular cognitive processes. However, a person's affective state may also influence the final selection.

For example, negative stimuli may distract the brain. Based on studies involving faces, it was determined that humans respond to facial stimuli specially. In particular, the brain will attend to faces depicting negative facial expressions. This effect was shown by an experiment in which participants had to decide whether faces in a grid of photographs were all the same or contained a discrepant (an angry or a happy face) (Hansen & Hansen, 1988). The participants seemed to pick out angry faces faster than they could pick out those showing happy expressions, which was thought to be a result of humans naturally attending to negative expressions first. Coined the "threat-superiority effect," the phenomenon in which humans have an early bias towards negative material in the surroundings was found to alter a person's current attentive state.

Because humans might naturally be attentive to anything that might be a threat as part of evolutionary survivability goals, it would make sense that threat-related stimuli are attended to immediately (Fox, 2008, p. 170). In this respect, a negative face or vocalization in the environment might draw a person's attention, thereby altering his attentive state.

Moreover, the brain may not attend the negative stimuli consciously, but rather subconsciously. Just the presence of the stimuli in a scene may influence brain processes even when they are not attended to. Many experiments have shown that the presence of a negative stimulus in an environment uniquely results in increased activity in the fusiform gyrus (for a fearful facial expression) and in the Superior Temporal Sulsus (for an angry voice). For example, in an experiment by Patrick Vuilleumier, a pair of faces and a pair of houses were aligned horizontally and vertically to form a cross with a central point of fixation. Participants needed to focus on the central point and match either the pair of faces or the pair of houses. When they matched the faces (and thereby attended to them), there was greater activity In the fusiform-gyrus region of the brain. However, just the presence of a fearful face enhanced activity in the fusiform-gyrus, regardless of whether the faces were attended to or not (Vuilleumier, Armony, Driver, & Dolan, 2001) (Fox, 2008, p. 175-176). Thus, the brain does not need to attend to negative stimuli in order to react to it. It responds directly to fearful stimuli in a scene, and may even weaken attention for that which was attended to initially. In general, the brain's bias toward negative stimuli (attended or not attended) may alter brain state at the moment of perception, which can subsequently influence how the environment is perceived.

Just doing it for the thrill

Furthermore, affect can influence how information, once perceived, is processed and can even select those details to be processed. There are two levels of processing a person's affective state may impact: general processing (the organization and interpretation of incoming stimuli) and judgment (the formation of opinions of stimuli for

evaluation).

The strategy used for processing information depends on if a person is in a positive mood or if he is in a negative mood. A person expressing happiness will assimilate stimuli from the environment into his own world view, connecting them to preconceived ideas of phenomena and further elaborating upon these stimuli (Fox, 2008, p. 215). The happy person will take in the sight of the mailman walking up his door steps and will know instantly that the mailman has come to deliver his daily mail. He knows that a mailman comes every day to deliver his mail and that every day he walks those same door steps up until the moment he slides the happy person's mail into the mail slot. Thus, the moment the mailman walks those steps again signals the happy person that he has received mail and that the mailman has come to deliver it.

While the happy man applies his own generalizations to the event, the more disturbed and fearful man will pay closer attention to the details of the event. He will apply less of his own ideas and attempt to accommodate them in light of the facts of his surroundings, carefully exhausting all aspects of stimuli until he can understand the environment and situation better (Fox, 2008, p. 215). Thus, he will instantly detect the mailman's torn pant pocket, his subtle limp, his loose shoelace, and his slight hesitance to lift the mail out of his mail bag. He will hypothesize that these details might all be related in some way to the mailman's unusual tardiness and may ponder what had happened to the mailman this day.

Both men differ in their moods, and thus, in their approaches to how they perceive the world. The more negative felt man will notice more stimuli than the happy man. It is thought that when a person is in a positive mood, his mind is at ease because there are no immediate obstacles to overcome or problems to solve. Thus, his mind flows and will readily explore the world and connect with it. However, a negative mood will come about when a person feels as if his present goals are threatened or have failed. The world does not seem consistent with his own view, so the person will abandon some of his notions in favor of the vast amounts of stimuli to be recorded (Fox, 2008, p. 215). Thus, both strategies differ in the level of details perceived.

Though, it is important to note that neither strategy is better than the other, for it depends on the situation. The strategy that allows the person to explore his surroundings more (the positive mood case) might be better in those situations in which creativity is required (e.g. finding the exit to a cave); on the other hand, when a more conservative approach is needed (e.g. shooting a ball into a basket), it might be more helpful if a person relied on the actuality of his situation (the negative mood case) (Gasper & Isbell, 2007, p. 97). It is possible that a person might even vacillate between positive and negative moods throughout the completion of a task in order to satisfy a need for either strategy.

The person's mood may also sway him towards particular choices within a task. In one experiment, participants were shown faces that varied in facial expression and then were subsequently asked within a gambling task to choose from a set of options that varied in risk. Those who saw positive facial expressions were more likely to choose risky options, while those who saw negative facial expressions tended to be more risk-averse (Winkielman & Trujillo, 2007, p. 81). The experimenters concluded that participants who were more fearful tended to make risk-averse choices whereas those who were sad would make more risky choices (to gain greater rewards). Thus, one's mood may determine the type of choice to take.

Memory

After the brain processes stimuli, it must then store them into memory for later recall. Affect may influence the memory of stimuli at either the stage of encoding (when stimuli are converted into forms appropriate for storage) or at the stage of consolidation (when memory of stimuli is stabilized for storage). During encoding, the amygdala will have increased activity when emotionally arousing stimuli is present. This increased activity is thought to lead to more improved memory for these stimuli and for events consisting of the stimuli (Fox, 2008, p. 196). For example, we might remember our own wedding day more vividly than any other event because we exhibited a wide spectrum of feelings on that day. Some attribute this increased memory for

emotionally arousing events to a priming effect in which we constantly replay, and therefore, rehearse these events in our mind (we want to relive the feeling of joy from our wedding day as much as possible). In any case, affect strengthens the recall for emotionally arousing events.

Affect may further determine the level of detail remembered from an event. As a side effect of increased attention and focus on emotionally arousing stimuli during the processing of stimuli, memory for other stimuli in a scene may suffer (Fox, 2008, p. 194). A woman's scream in a park might divert one's attention, and thus, he might not have any recollection of the passerby's wardrobe or the type of fish in the pond. Additionally, a person's affective state might simply tag details of an event for later recall. The phenomenon known as the "mood congruency effect" occurs when people remember an event better if they express the same affective state as they were in during the event (Fox, 2008, p. 212). During a party, a person in a joyful mood may perceive the brands of wrist watches belonging to the people with whom he interacts, but may later recall those brands better within the same joyous state. One's affective state may regulate the depth of information recollection, either strengthening details or forming gaps in the knowledge of events.

During consolidation, affect can manipulate our judgment regarding the quality of stored emotionally arousing information from the encoding stage. In particular, it has been shown that there is a high correlation between the vividness of the memory of an event and the amount of emotion displayed or experienced during that event (Fox, 2008, p. 192). For example, many who witnessed the Challenger catastrophe could recall all sorts of details such as who was sitting next to them as they watched the account and how the newscaster delivered the announcement. However, not all these details were necessarily accurate. As shown in one particular study, researchers compared students' reports of how they heard about the news of the Challenger disaster the morning after the event to their reports three years later. They found great discrepancy between the reports, with students distorting various details such as where they had actually heard the news. The brain simply filled in these facts and successfully convinced the students that they were valid. Thus, affect might heighten our confi-

dence in our recollection of an emotionally arousing event, effectively masking any inaccuracies in our account.

To act or not to act?

When the decision has been made to take on a new task, a person's affective state can influence both whether he should carry out the task and how well he will perform on the task. Specifically, a person will move through two modes when preparing for the task: task assessment and competency assessment. The person will begin in task assessment where he will attempt to acquire as much information about the task such as what the task entails. After understanding the task, he will move into the mode of competency assessment in which he takes inventory of himself, wondering if he has the skills and aptitude necessary to carry out the task (Gasper & Isbell, 2007, p. 105). For example, a person, upon understanding that in order to save the kitten from the tree he would have to climb it, would wonder if he is physically fit to climb and whether he possesses the level of determination required to make it to the top.

A person's mood can promote either task assessment or competency assessment. It was found that those in sad moods tended to stay more in the task assessment mode. As noted earlier, a person in a negative mood will perceive more stimuli and will be more cautious about details he encounters; thus, he will strive for a more detailed understanding of the task before he proceeds to competency assessment. However, the sad person may not fare well in competency assessment, for he may not take up a task due to inhibitions or pessimism. On the other hand, a positive person would be more optimistic about carrying out the task. He would be more willing to look past any negatives such as a lack of skills and may be more permissive to task failure, attempting to improve his skills upon any obstacles. However, the positive person might not exhaust all possible details like the negative person, and may prematurely jump to a competency assessment with only a partial understanding of the task (Gasper & Isbell, 2007, p. 105). Thus, the person's mood can affect both his desire and his ability to perform a task.

At a higher level, a person's mood can influence how productive he is in carrying out a task. Mihaly Csikszentmihalyi, a researcher in positive psychology, developed the notion of "flow," a state of concentration in a task in which the person exhibits a high level of focus and attention (Figure 6). In "flow," a person acts automatically throughout the duration of the task such that time feels unmoved, almost as if he were in a trance. He absorbs the task completely, bouncing off of feedback the environment gives in response to his individual actions.

In order for one to achieve a state of flow, the task must be highly challenging and one must have the right amount of skill. If the task is too difficult such that one does not have an adequate skill set, one will become anxious. Conversely, an incredibly easy task might simply bore the moderately skilled. Furthermore, the task completer needs to be able to overlook his own shortcomings and persevere through any obstacles, feats only achievable in a positive mood and not in a negative one. Depression may hinder one's ability to see the intermediate goals that comprise a task and fear may simply discourage him in every seemingly straggling moment (Csikszentmihalyi, 1990). Thus, one can only attain flow by enlisting in a positive emotional state.

At the same time, flow can induce a positive mood. One of Csikszentmihalyi main motivations for his concept of flow is that he believes a state of flow in itself leads to satisfaction and happiness, for it provides a sense of accomplishment and meaning within individuals. People will set goals for themselves, solely for the purpose of achieving them and obtaining some reward. Furthermore, it provides a sense of control over a task, a feeling that everything unfolds as planned and understood from experience – an environment conducive to one who will ex-

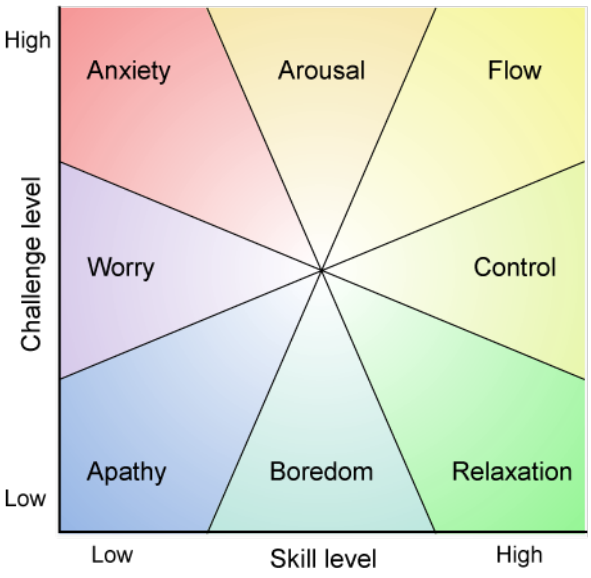


Figure 6: Mental state in terms of challenge level and skill level as defined by Csikszentmihalyi
http://en.wikipedia.org/wiki/File:Challenge_vs_skill.svg

hibit a positive mood (Csikszentmihalyi, 1990). Thus, a happier emotional state may lead to flow, which may then generate further happiness.

One's affective state may determine how one sees and behaves in the world. It can narrow the set of choices from which one might choose and it can lead him down one particular path. And it does not influence just momentary decisions – it can define future perception of past events and stimuli. However, it is important to note that while the affective state of a person may be detrimental to him in certain scenarios, it is the situation itself that decides whether one's mood is leading him in the right path or merely straying him from the correct one. Thus, there is much interplay between affect and the demands of cognition; our emotional state might prompt us to behave in a certain way, but the turn of a situation may force upon a different state that may either help or hinder our efforts. It is then our determination to change our outlook, and thereby our affective state, that brings us back on track in accomplishing our various goals.

3 Ekman and the Facial Action Coding System (FACS)

The Basic Emotions

Paul Ekman, a renowned researcher in human emotions and facial expressions, had identified seven basic human emotions that were universally expressed and recognized. These seven were happiness, anger, sadness, disgust, fear, surprise, and contempt. He proved the universality of the first six emotions through a series of facial expression-emotion matching experiments. In 1969, Ekman and his colleagues showed a set of photographs of males and females to students in both a college in the United States and a college in Brazil (Ekman & Friesen, 1969). The students had to pick from a list of eight affect terms an emotion that represented the depicted facial expression. The researchers found that within each group of students, there was high agreement on these interpretations for the first six emotions, and that their own la-

beling matched the majority of that of the students. Furthermore, the interpretations were almost the same across both cultures. From these results, they concluded that these particular emotions were expressed and recognized universally.

Subsequent experiments sought to strengthen their findings. They had suspected that Americans and Brazilians could have acquired similar interpretations for emotions through the same sources of mass media. If this fact were so, then the two cultures would not be truly independent of each other, weakening a cross-cultural claim. To remove the effect of shared mass media, Ekman and his colleagues went to the highlands of southeast New Guinea to conduct the same experiments with a culture that was much isolated from the media and Western civilization (Ekman & Friesen, 1969). The results of this experiment were consistent with those in their first experiment, showing that indeed the emotions were universal. Moreover, Ekman showed that not only was the judgment of affect universal but also the actual expression. In a later study, he found that Americans and Japanese produced the same facial expressions (matching the ones used in previous studies) through the same facial muscles as they individually watched an emotionally arousing film (Ekman, 1972). Thus, with these experiments and several later ones, he could provide enough support for the universality of the six emotions.

He later showed that the seventh emotion, contempt, was also universal through similar experiments with people from West Sumatra, Indonesia (Ekman & Heider, 1988).

See Table 4 for a full list of the seven basic emotions.

Table 4: The seven basic emotions.

Emotion
Happiness (Joy)
Sadness
Surprise
Fear
Anger
Disgust
Contempt

Facial Action Units

Paul Ekman and colleague Wallace Friesen later developed the Facial Action Coding System (FACS), a systematic method for distinguishing the facial expressions depicting the basic emotions (Ekman & Friesen, 1978). They identified 44 Action Units (Table 5), the independent movements of individual facial muscles, that could summarize all possible facial expressions. They then matched subsets of these Action Units to the specific facial expressions of the basic emotions (Table 7). For example, for a facial expression of happiness, the Action Units 6 and 12 (Raising of the Cheek and Pulling of Lip Corners) are active (Ekman & Friesen, 1978).

Table 5: The Action Unit (AU) codes describing various facial muscles. (Ekman & Friesen, 1978)

AU No.	FACS Name	Muscular Basis
1	Inner Brow Raiser	Frontalis, Pars Medialis
2	Outer Brow Raiser	Frontalis, Pars Lateralis
4	Brow Lowerer	Depressor Glabellae; Depressor Supercilli; Corrugator
5	Upper Lid Raiser	Levator Palebrae Superioris
6	Cheek Raiser	Orbicularis Oculi, Pars Orbitalis
7	Lid Tightener	Orbicularis Oculi, Pars Palebralis
8	Lips Toward	Orbicularis Oris Each Other
9	Nose Wrinkler	Levator Labii Superioris, Alaeque Nasi
10	Upper Lip Raiser	Levator Labii Superioris, Caput Infraorbitalis
11	Nasolabial Furrow	Zygomatic Minor Deepener
12	Lip Corner Puller	Zygomatic Major
13	Cheek puffer	Caninus
14	Dimpler	Buccinnator
15	Lip Corner Depressor	Triangularis
16	Lower Lip Depressor	Depressor Labii
17	Chin Raiser	Mentalis
18	Lip Puckerer	Incisivii Labii Superioris; Incisivii Labii Inferioris
19	Tongue Show	
20	Lip Stretcher	Risorius
21	Neck Tightener	platysma
22	Lip Funneler	Orbicularis Oris
23	Lip Tightner	Orbicularis Oris
24	Lip Pressor	Orbicularis Oris
25	Lips Part	Depressor Labii, or Relaxation of Mentalis or Orbicularis Oris

Table 6: The Action Unit (AU) codes describing various facial muscles. (continued) (Ekman & Friesen, 1978)

AU No.	FACS Name	Muscular Basis
26	Jaw Drop	Masetter; Temporal and Internal Pterygoid
27	Mouth Stretch	Ptergoids; Digastric
28	Lip suck	Orbicularis Oris
29	Jaw Thrust	
30	Jaw Sideways	
31	Jaw Clencher	masseter
32	[Lip] Bite	
33	[Cheek] Blow	
34	[Cheek] Puff	
35	[Cheek] Suck	
36	[Tongue] Bulge	
37	Lip Wipe	
38	Nostril Dilator	Nasalis, Pars Alaris
39	Nostril Compressor	Nasalis, Pars Transversa and Depressor Septi Nasi
41	Lid Droop	Relaxation of Levator Palpebrae Superioris
42	Slit	Orbicularis Oculi
43	Eyes Closed	Relaxation of Levator Palpebrae Superioris
44	Squint	Orbicularis Oculi, Pars Palpebralis
45	Blink	Relaxation of Levator Palpebrae and Contraction o Orbicularis oculi, Pars Palpebralis
46	Wink	Orbicularis Oculi

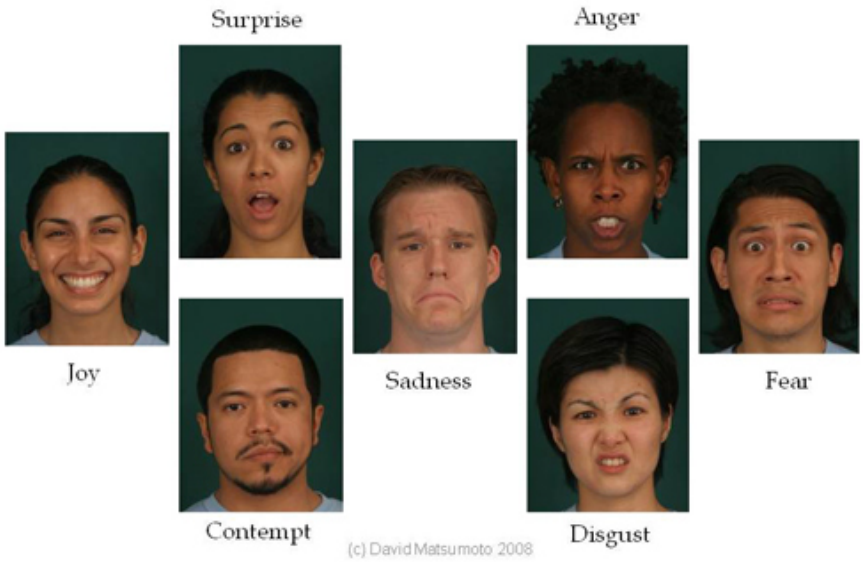


Figure 7: The seven basic emotions as facial expressions.
David Matsumuto http://www.apa.org/science/about/psa/2011/05/facial-expressions.aspx

Table 7: Emotional coding based on action units (AUs). A subject displays an emotion when its AUs are present. (Ekman & Friesen, 1978)

Emotion	Action Units
Happiness	6+12
Sadness	1+4+15
Surprise	1+2+5B+26
Fear	1+2+4+5+20+26
Anger	4+5+7+23
Disgust	9+15+16
Contempt	R12A+R14A

Part III

Facial Expression Recognition

Overview

First a face detector detects a face in a given image using a pre-trained face model. The detector "fits" the face model over the image by scaling and warping certain features of the model until it matches a face in the image. The parameters of the fitted model are then input into a trained facial expression classifier that will output a classification label (happy, sad, angry, etc). Figure 8 describes the high level process.

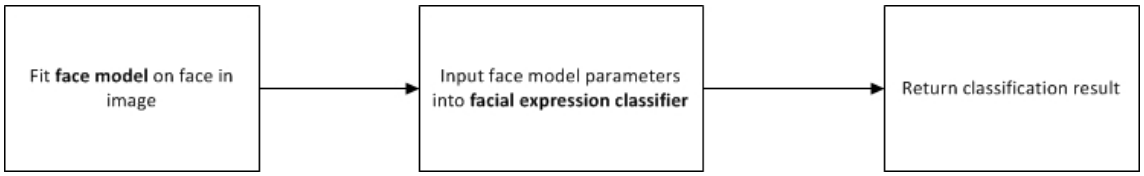


Figure 8: Facial expression recognition high level process

4 Face Model Training

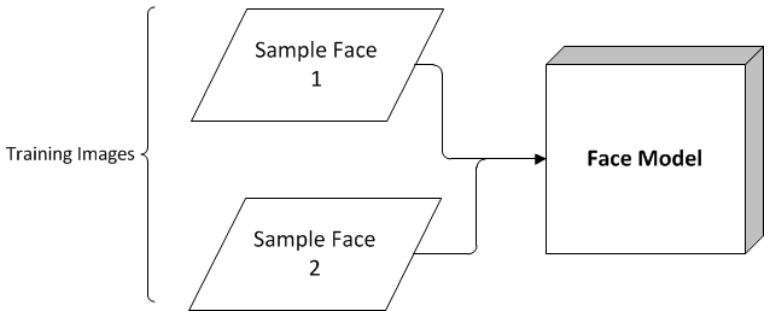


Figure 9: Training the face model to recognize human faces

4.1 Active Appearance Models

A deformable face model is constructed from training with a set of face images that have been annotated with landmark points, which define the main features of the face (Figure 9). The same number of landmarks are used in each image and are placed on the same set of facial features. Using Procrustes Analysis, in which one object is scaled, translated, rotated, and reflected to coincide with another, these landmarks are aligned

across the training images to create a statistical shape model. The training images are then warped to fit the mean shape of the shape model (by aligning the landmark points in each image to those of the mean shape) and placed into a normalized texture vector. A statistical texture model is then built from this texture vector using some sort of eigenvector analysis such as Principal Component Analysis (PCA) (Cootes, Edwards, & Taylor, 2001).

Principal Component Analysis (PCA) is used to convert the sequences of landmark points for each training image into normalized eigenvectors and eigenvalues. A face can then be described in terms of the linear combination of these eigenvectors:

$$X = \overline{X} + \sum_{i=1}^{n} p_i b_i$$

or

$$X = \overline{X} + Pb$$

where $\overline{X}$ is the "mean-shape" observed during PCA, P is the set of PCA eigenvectors, and b is a vector of parameters for each eigenvector. This is known as the point distribution model (PDM) of the Active Appearance Model (Cootes et al., 2001).

5 Face Model Fitting

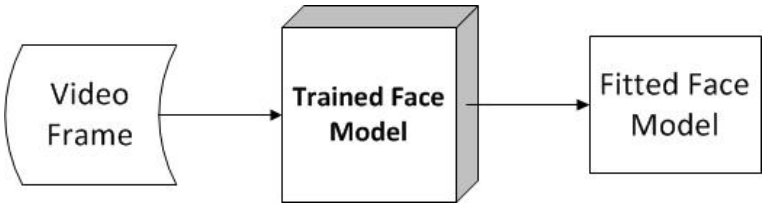


Figure 10: Fitting a face model onto a face in a target video frame.

A trained deformable model is then "fitted" on a face in a source image (Figure 10). The fitting process requires determining the parameters of the deformable model such that the model's landmarks correspond to the same matching regions of the image. A shape X will be generated from the shape model with a set of parameters b. This

shape will be placed on a region of the image, and the pixels within this region will be sampled, adjusted to the texture model's frame, and then compared to the texture G generated from the texture model with the same set of parameters b. The error between G and the adjusted sample will drive changes in the b parameters for the next iteration of the search. The model is fitted on the face in the image once this error converges at a value (Figure 11).

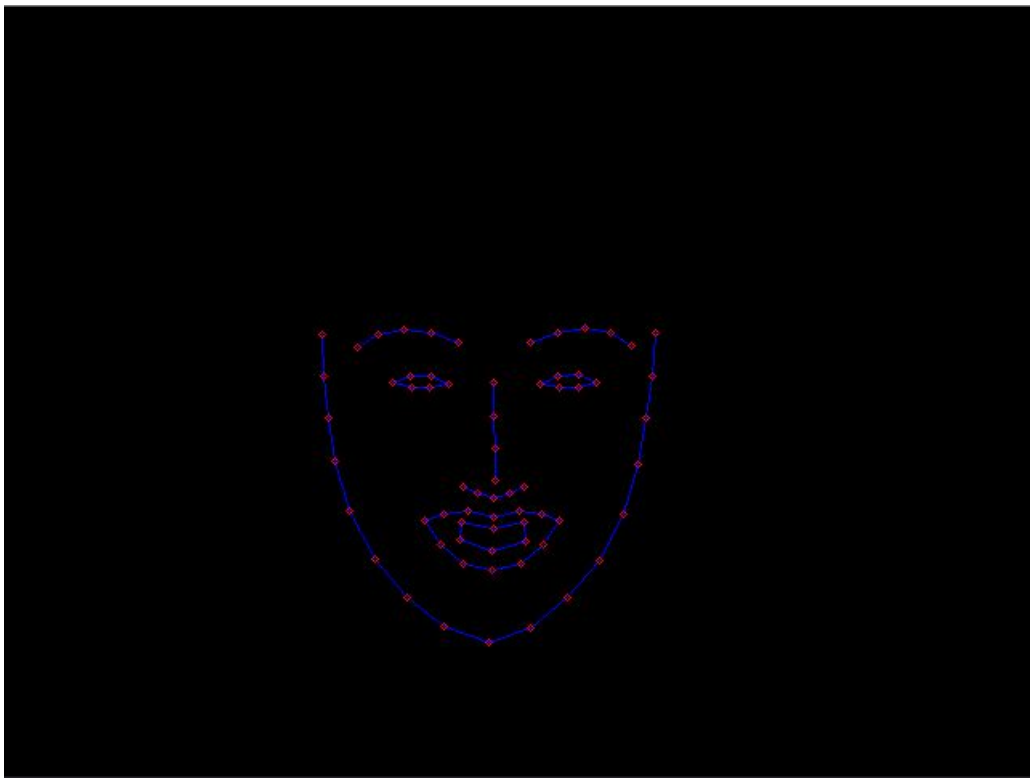


Figure 11: Face model with 66 vertices constructed from Jason Saragih's FaceTracker.

6 Face Model Expression Classification

Supervised classification can predict the type or identity of an unseen object. It takes as input a set of numerical values describing the target object and outputs a set of classification labels with their likelihoods of matching the object. There are many supervised classifier algorithms and approaches (to name a few, decision trees, the nearest neighbor algorithm, and artificial neural networks (ANN)). In the experiment, we use support vector machines (SVM) for predicting classification labels (angry, happy, etc)

for facial models.

Support Vector Machines

Support Vector Machines (SVM) classify input observations as belonging to one of two classes: a positive class or a negative class (an extension to the SVM is the multiclass-SVM, which provides support for more than two classes. The SVM aims to maximize the distance between a hyper plane and the two classes it separates, constructing an optimal hyper plane that defines the two classes. Thus, for an input described as some vector x, a trained SVM will output

$$w^T x + b \geq +1 \textbf{ or } w^T x + b \leq -1$$

for x belonging to either the positive or negative class. W is the linear sum of several support vectors that, together with a constant b, define the optimal hyper plane. The support vectors are those training vectors which define the edge between their class and the alternative.

Thus, for support vectors

$$[(x_1, y_1), (x_i, y_i), .., (x_n, y_n)] \textbf{ st. } y_i \in \{+1, -1\}$$

we have that

$$(w^T x_i + b) = y_i \in \{+1, -1\}$$

or simply

$$y_i(w^T x_i + b) = 1$$

If we let h be the distance between one support vector (x_i, y_i) and the optimal hyper plane, then

$$h = \frac{y_i \in \{+1, -1\}}{\|w\|}$$

To maximize this distance, SVMs during training seek to minimize w through optimization techniques.

SVMs in Facial Expression

Recall that to describe a particular face X in the Active Appearance Model, we have

$$X = \overline{X} + Pb$$

where $\overline{X}$ is the mean face shape, P is a set of eigenvectors describing various facial features, and b is a vector of parameters that manipulate the facial features. Then, b is essentially the **signature** for a particular facial expression for a specific facial perspective. The vector b is further composed of values which describe the facial orientation (global scaling, rotation, and translation) and the manipulation of the frontal facial features. Thus,

$$b = \{s, r, t, c\}$$

where s is the global scaling parameter, r is the rotation parameter, t is the translation parameter, and c is the frontal facial feature parameter. The frontal facial feature parameter describes the normalized face (after scale, rotation, and translation are fixed), and thus, this parameter is the only one significant in describing the actual expression. We can train our SVM using the parameter c, and subsequently classify new observed facial expressions through the vector constructed by c (after constructing the vector b).

In particular, a dataset is generated consisting of different vectors c labeled with their respective facial expressions (in SVM, the labels must be integers, so we map each facial expression word to an integer label). An SVM model is trained from this dataset, and this model is used to predict expression classifications for new observations (described by a vector c).

Part IV

Effectiveness in Application

7 AutoTutor

AutoTutor is an intelligent tutoring system (ITS) that aims to teach and assist students in various topics such as physics and computer literacy. Through the use of natural language, AutoTutor attempts to broaden a user's understanding of a topic by helping him construct an explanation for a phenomenon. AutoTutor will first pose a question within the topic. The user will reply to this question by typing his response in a textbox provided on the screen. Based on this response, the system will make various dialog moves. It may provide feedback to the user, "pump" for a more detailed answer, give hints, make a claim that might be useful in getting to the final answer, or summarize what has been discussed (D'Mello et al., 2008). Like a real tutor, AutoTutor attempts to engage the learner beyond simple fact-recall in an attempt to truly achieve an understanding of a topic.

Recently, work has been done to create an affect-sensitive AutoTutor, a version of the system that would regulate its dialogue moves according to the user's affective states. Based on prior research, the researchers behind the system had identified six relevant affective states in learning that they would attempt to monitor: boredom, engagement/flow, surprise, delight, frustration, and confusion (D'Mello et al., 2008). The system would detect these states through various sensors and devices. Facial recognition was performed through an IBM BlueEyes camera based system that can track a subject's face in real time, decompose the face into action units, and produce a final emotion classification. Posture detection was also employed through the use of a Body Posture Measurement System (BPMS), a pressure pad that can be placed on any surface (such as on a user's chair seat and chair back). The pad measures the amount of pressure exerted on it at various regions. It was thought that increased pressure on either the seat or the back would correspond to the user leaning forward (expressing a high level of attentiveness) or sitting back (a low level of attentiveness).

The researchers had reduced the initial set of affective states to detect to just three: confusion, boredom, and frustration. They had discovered in their prior research that delight and surprise emotions were expressed very rarely in learners, and they did

Student Model	Tutor Action
Current Emotion	**Feedback**
boredom, confusion, frustration	positive, neutral, negative
Classification Confidence	**Empathetic and motivation statement**
high or low	
Previous Emotion	**Next Dialogue Move**
boredom, confusion, frustration	hint, pump, prompt, splice, assertion
Global Student Ability	**Facial Expression**
high or low	surprise, delight, compassion, skeptical
Quality of Current Answer	**Speech Information**
high or low	pitch, intensity, speech rate, etc

Table 8: AutoTutor rules that acted on the user's current affecive and learning states (D'Mello et al., 2008)

not want to risk interrupting through a change in the system's behavior any state of engagement/flow detected in the user. To respond to the affective states that were detected, a set of rules were derived (Table 8) that took into account a user's current/previous emotional state, overall performance, quality of recent response, and confidence level (D'Mello et al., 2008). The output of each rule would be the tone of feedback given (positive, negative, neutral), some empathetic/motivational statement (to alleviate negative affective states such as frustration), a dialogue move, a facial expression (delivered through a virtual agent, meant for displaying compassion), and the speech intonation of AutoTutor's vocal responses (also delivered through the virtual agent) (D'Mello et al., 2008).

Their specific method for integrating their detection mechanisms was to use a multilevel classifier in which classifications of the raw detected data were performed individually by each sensor (or sensor group) first and then a final classification was made of these individual classifications using a super classifier.

8 My Affect-sensitive Game System

Overview

We designed and developed our own affect-mediated system to further explore the benefits of affective computing. The system is centered around a game known as the BasketGame. It consists of three parts: a facial expression recognition engine, a client middleman component (DetectClient), and the game itself. The facial expression recognition engine is a modified version of Jason Saragih's FaceTracker (Saragih, S.Lucey, & Cohn, September, 2009). The BasketGame was designed from scratch and uses DetectClient to communicate with the recognition engine. (Figure 12)

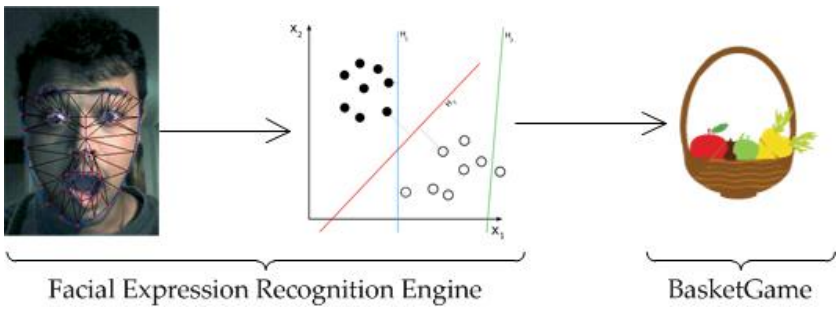


Figure 12: A high level overview of the system. The Facial Expression Recognition Engine processes video frames from a web camera, tracks the face in each frame, and classifies the face as expressing one of the seven basic emotions (or the neutral expression). It then sends this emotion label to the BasketGame to be used for manipulating the game's logic.

Facial Expression Recognition Engine

Technical Specifications

The default FaceTracker application developed by Jason Saragih was built using C/C++ and OpenCV 2.x (Saragih et al., September, 2009). The modified application uses WinSock2. In total, there are 3108 lines of C/C++ code.

Overview

The Recognition engine is based off of Jason Saragih's FaceTracker application, which uses an improvement to the traditional Constrained Local Model (CLM), a name Saragih gives to the class of face fitting approaches that includes the Active Appearance Model (Cootes et al., 2001). This improved CLM performs better than the other CLMs for generalized face alignment of unseen subjects (Face Alignment through subspace). FaceTracker performs face alignment and tracking using models that were pre-trained with images from CMU's MultiPIE face database (Gross, Matthews, Cohn, Kanade, & Baker, 2008, 2009). The program constructs a mesh consisting of 66 points (x,y) on top of a face in the image and deforms the mesh such that the points line up with their assigned regions (determined during face model training) (Figure 13).

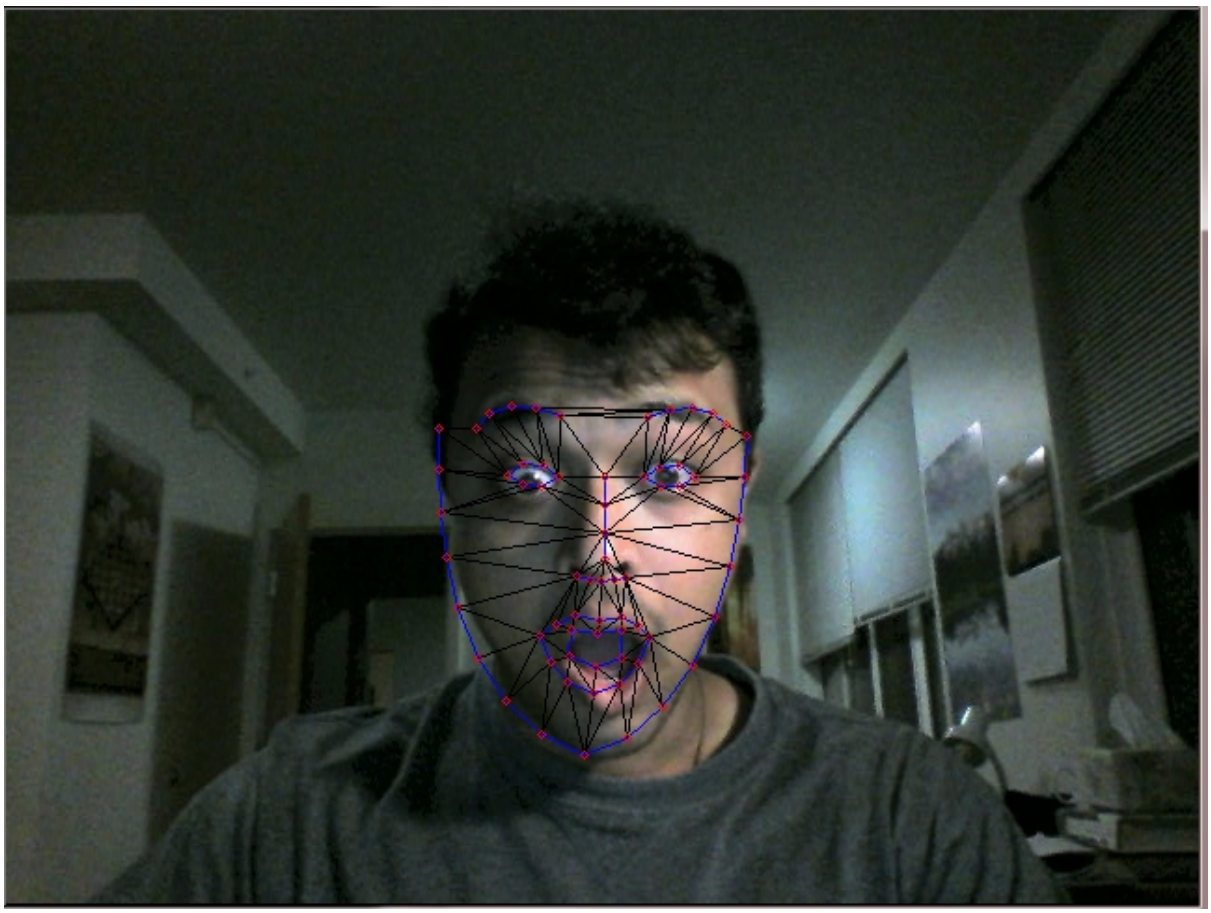


Figure 13: Jason Saragih's FaceTracker fitting a face AAM on a target face

We extended FaceTracker to provide support for facial expression recognition. The

modified application uses support vector machines (SVM) to classify the seven basic facial expressions and a neutral (emotion-less) expression in real-time from webcam video (Figure 14). LIBSVM was used to train the SVM model with a set of images from the Extended Cohn-Kanade (CK+) dataset (Chang & Lin, 2011) (Lucey et al., 2010). SVM prediction code from LIBSVM was embedded directly into the engine to allow for classification of facial model parameters generated by FaceTracker's alignment code. Furthermore, to enable interoperability between the Recognition engine and applications that wish to consume it (such as a GUI developed in a managed environment), the application has a multithreaded server that listens and communicates with client applications on a socket. Clients can request to be updated on the current expression classification through a connection with the server.

Figure 14: Classification of FaceTracker models

The Extended Cohn-Kanade (CK+) Dataset

The Extended Cohn-Kanade (CK+) dataset is an improvement over its predecessor, the Cohn-Kanade (CK) dataset, a popular dataset used for detecting facial expressions.

The initial CK dataset consisted of several facial images that were FACS coded for Action Units (AU) and these AU labels were subsequently validated (Kanade, Cohn, & Tian, 2000). The new CK+ dataset is much larger than the CK dataset and contains validated emotion labels for a subset of dataset images in addition to validated AU labels for the entirety of the dataset (Lucey et al., 2010).

In our training dataset, we used solely the images that were emotion labeled as they expressed the correct emotion according to the FACS Investigator Guide (Ekman, Friesen, & Hager, 2002). The images were labeled with exactly one of the seven basic emotions (anger, contempt, disgust, fear, happiness, sadness, surprise). They were each part of a sequence of frames in which a subject transitioned from a neutral face to the target emotion (Figure 15). We used some of the initial neutral frames to represent the neutral emotion (lack of emotion) in our training dataset.

In their paper, Lucey et al trained an SVM with the emotion labeled data and evaluated its classification performance. They found that the SVM had the best performance classifying expressions of happiness, surprise, and disgust with prediction accuracy of 98.4%, 100%, and 68.4% respectively when only the AAM PDM parameters were classified. The classifications for the rest of the emotions were poor: 35% for anger, 21.7% for fear, 4% for sadness, and 25% for contempt. They explain these results by noting that happiness, surprise, and disgust involve large facial movements and more importantly move along the facial mesh constructed by the AAM. The other facial expressions are more subtle, and thus, may easily be confused with the stronger expressions. However, they do report greater prediction accuracy for these emotions when both the shape and the appearance (texture) of the

Figure 15: A sample training image from the CK+ dataset. This particular subject is expressing surprise.
©Jeffrey Cohn

model are classified: anger (75%), fear (65.2%), sadness (68%), and contempt (84.4%) (Lucey et al., 2010).

LIBSVM Training

The model used for facial expression prediction was trained with the CK+ emotion images using the LIBSVM toolkit (Chang & Lin, 2011). We used the easy.py script that came with the library to automate the scaling and parameter selection process during SVM training. The script uses an RBF kernel and performs 5-fold cross validation.The resulting model had a prediction accuracy of 70.59%. It seemed to work well in informal tests to predict expressions of happiness, surprise, disgust, and anger.

The training data that was input into the toolkit was generated by running our training dataset (consisting of CK+ images) into the Expression engine and logging the generated model parameters to a file (for simplicity, we only logged the shape parameters and not the appearance of the model). The parameters from all training images were labeled (a number from 1 to 8) and then consolidated into a single training file, which the easy.py command took in as input.

Interoperability

To pass Facial Recognition Engine results to other applications without requiring a direct reference to the engine (which may not be possible in a managed environment such as in Java or .NET), sockets are used to allow for inter-process communication capability. Specifically, we constructed a multithreaded server that listens for clients on a socket. Upon a client connection, the client may ask the server to send it messages on a consistent basis regarding the currently displayed emotion (Figure 16).

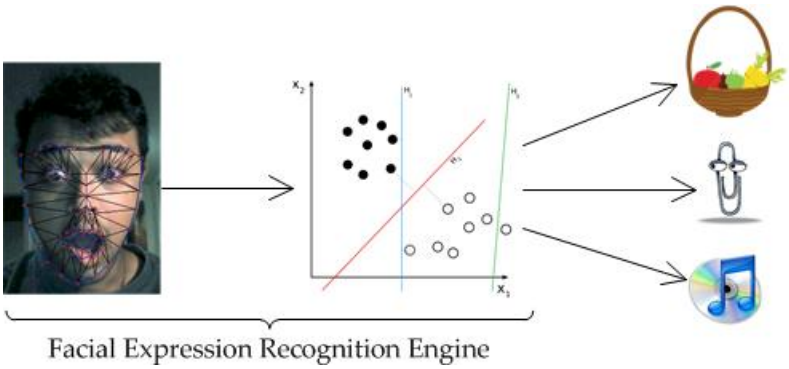


Figure 16: Any client can communicate with the Engine to acquire real time updates on the user's current emotional state.

DetectClient (client middleman)

Technical Specifications

DetectClient is written in C# and was designed to be referenced in a .NET project as DetectClient.dll. It has 71 lines of code across 2 classes.

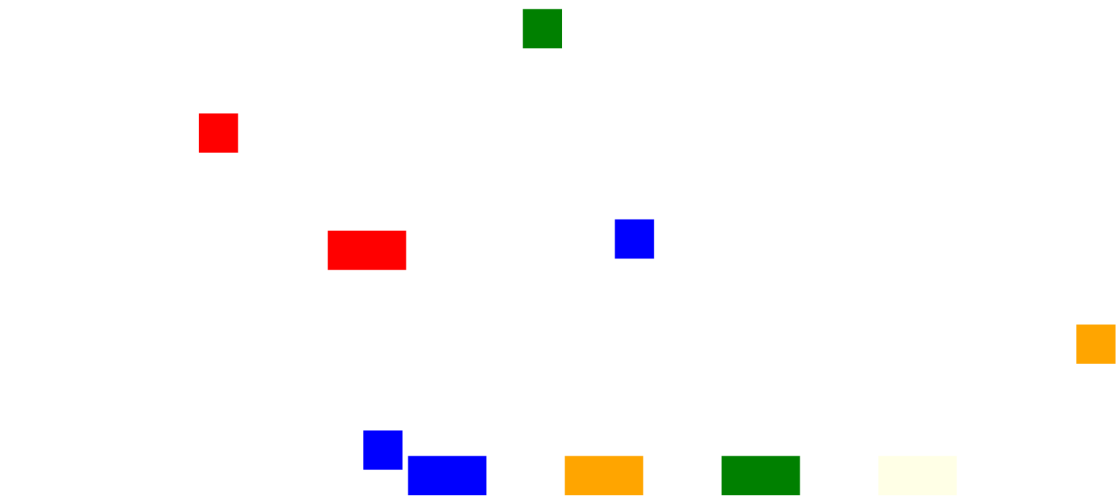
Overview

The purpose of DetectClient is to abstract away from a consuming application the necessary connection backbone with the Facial Expression Recognition Engine. It sets up a connection with the Engine and requests for constant notifications of the current facial expression. For every emotion message received from the server, DetectClient will raise an EmotionChanged event, which can be subscribed to by the consuming application. Thus, in order for a .NET application to be emotionally aware, it only needs to instantiate the DetectClient and subscribe to the EmotionChanged event. Then, it is free to handle emotion data however it pleases.

The BasketGame

Technical Specifications

The BasketGame is written in C# and utilizes the Windows Presentation Foundation (WPF) framework to render the GUI. It uses the Model-View-ViewModel (MVVM) design pattern. It has 552 lines of code across 32 classes.



N

Figure 17: Early game skeleton used to adjust game logic and mechanics. The rectangles represent the ”baskets” that players would drag around the screen. Squares represent the falling items to catch.

Overview

The BasketGame was designed for young children, specifically those that would be tested in the actual experiment. In our initial game planning, we had wanted to construct a game that had elements that could change without sacrificing usability or interrupting the participant's mental model. We felt that a game which did not initially display all possible objects for interaction, but only revealed those objects one by one

Figure 18: Final BasketGame snapshot

would be perfect for an adaptive game system. We could twist the rules for how those objects appeared while still meeting the expectations of our users.

Gameplay

Food pieces of various colors will fall from the sky. There are different colored baskets that match the colors of the food. The goal of the game is to catch the food with the same color basket. We designed two themes for the game, a fruit theme (Figure 18) and a vegetable theme (Figure 19).

The game is divided into phases (levels). A phase describes certain parameters of the game: the variety of colors of food that can drop, the number of different locations food items will drop, and the rate at which each item falls (Table 9). Based on the player's accuracy in catching the food in matching baskets in a phase, the game will move onto the next phase in which a greater variety of food will drop in more random locations, possibly at an increased rate. If a player performs poorly in a phase, however, the game will move back to the previous, easier phase.

Figure 19: A second theme for the game using vegetables instead of fruits.

Game logic

We created two game engines: the Simple Game Engine and the Affect-mediated Game Engine, the affect-sensitive component of the Game System. Both game engines can easily be swapped out for the other.

Simple Game Engine

The game changes the currently loaded phase based on the player's performance. A player's performance is tracked utilizing two running counts: a positive streak and a negative streak. We increment the positive streak whenever the player catches a food item and increment the negative streak for every miss of an item. The two streak counts oppose each other: every time the positive streak increases, the negative streak decreases, and vice versa. If the positive streak meets the global max streak threshold, the game progresses into the next phase (a harder phase). However, if the negative streak reaches the max streak threshold, the game loads the previous phase (the easier phase). The streaks are reset for every load of a phase. The intent of the streak system is to model the player's consistent, stable performance, while accounting for

Table 9: The five levels of the game (final study settings). A level is described by its speed (how fast items drop), its colors (the number of different colored items that spawn), and its spawn locations (the number of different locations on the screen where the items will drop)

Level	Speed	Colors	Spawn Locations
1	slow	2	2
2	slow	3	10
3	medium	4	10
4	medium	5	10
5	fast	5	15

deviations that would be acceptable in our concept of "good" or "poor" performance (for example, a miss within a sequence of ten successful catches would still be a good streak since the player caught the majority of items).

The concept of winning is defined by a running score: when the score reaches a max score threshold, the player wins the game. The score number is not directly visible to the player as we did not want to assume that all our participants could read numbers. Furthermore, we did not want those participants who would be able to read and understand the score number to skew our results because of some additional effect. Thus, we kept the score number hidden and visibly expressed the score through a progress bar at the bottom of the game screen.

Specifically, this progress bar consists of a graphic of a basket that will move across the bottom from left to right for changes in the score number. It will move one step to the right for a score increment and one step to the left for a score decrement. When the score number reaches a max score threshold, indicating the end of the game, the basket will be at the far right, touching the graphic of a star. Participants look to this progress bar for information regarding how close they are to winning the game.

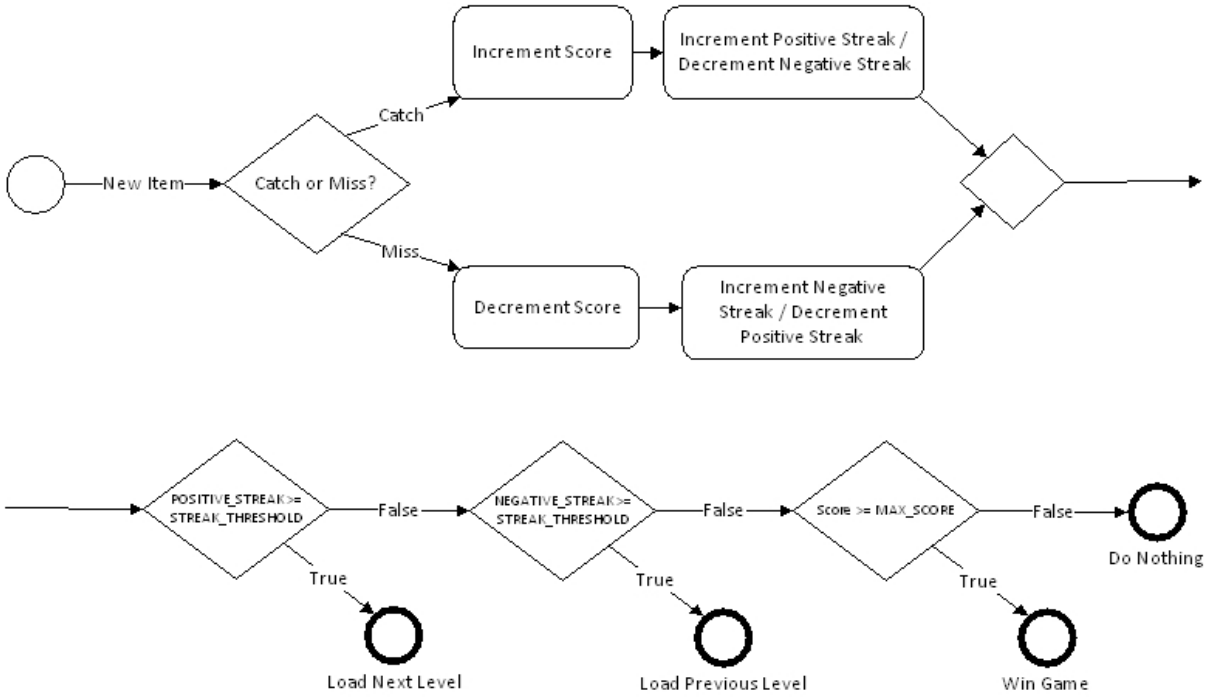


Figure 20: Default (control) game logic to handle every item event (the catch or miss of a spawned item)

Affect-mediated Game Engine

The Affect-mediated Game Engine is an extension of the Simple Game Engine in that it additionally reacts to negative emotions. We split up the basic emotions and a neutral emotion classification into two sets: positive and negative. The positive emotions are: happiness, content, and the neutral state of no emotion. The negative emotions are the remaining five emotions: fear, disgust, surprise, anger, and sadness. It is important to note that we define a negative emotion as one that is expressed in the course of struggle during the duration of a task, as some of these include emotions we had observed from children who were frustrated during an activity. Conversely, emotions in the positive set, such as happiness and neutral, were observed from children who did not express any frustration or internal struggle during their activities.

The negative emotions directly influence the game logic. Whenever a player exhibits a negative emotion, the previous game phase is **immediately** loaded, regardless of the player's positive or negative streak. The game does not adapt or accommodate positive emotions; in the case that a player exhibits an emotion from the positive set,

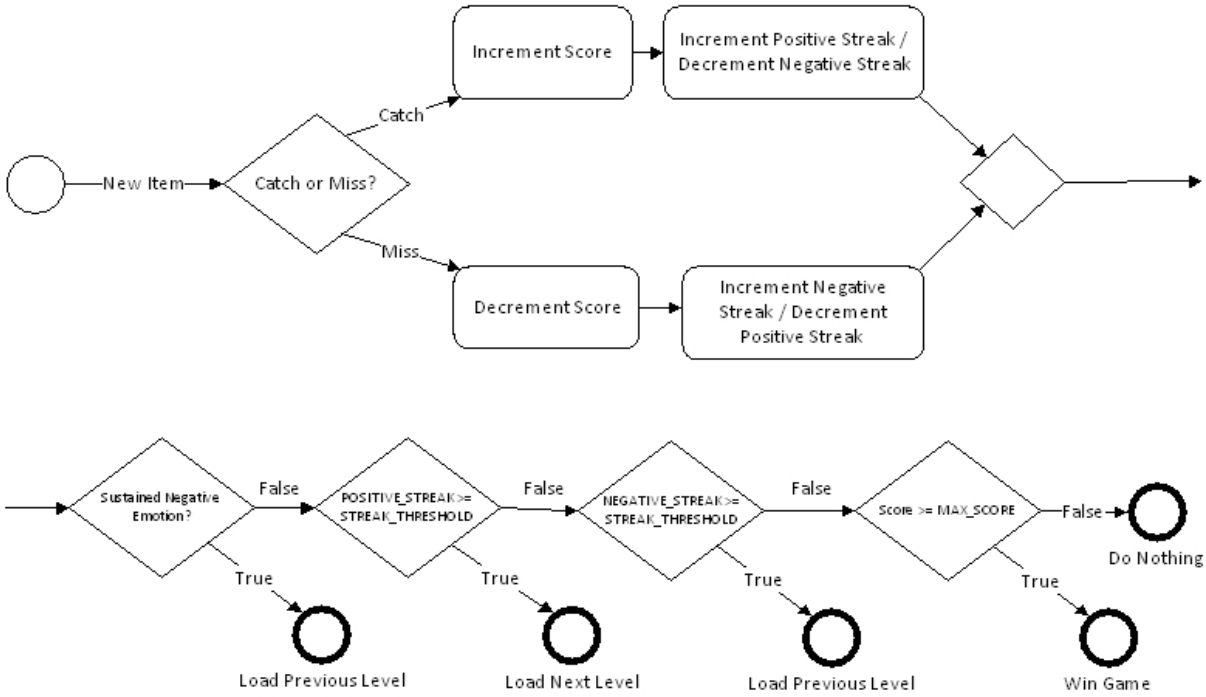


Figure 21: Affect game engine logic. Note: a sustained negative emotion will immediately load the previous level

the game will follow the default logic from the Simple Game Engine, in which phase advances and regressions depend on the player's positive and negative streaks.

The difference between the two game engines can be seen in Figure 22. The user misses a few items and begins to show surprise. With the Simple Game Engine, the game will load the previous level only after the user misses twelve items, which is the global max threshold value. However, the Affect-mediated Game Engine will load the previous level at the first detection of sustained surprise.

An example of the data that is collected by the game can be found in Appendix 13.

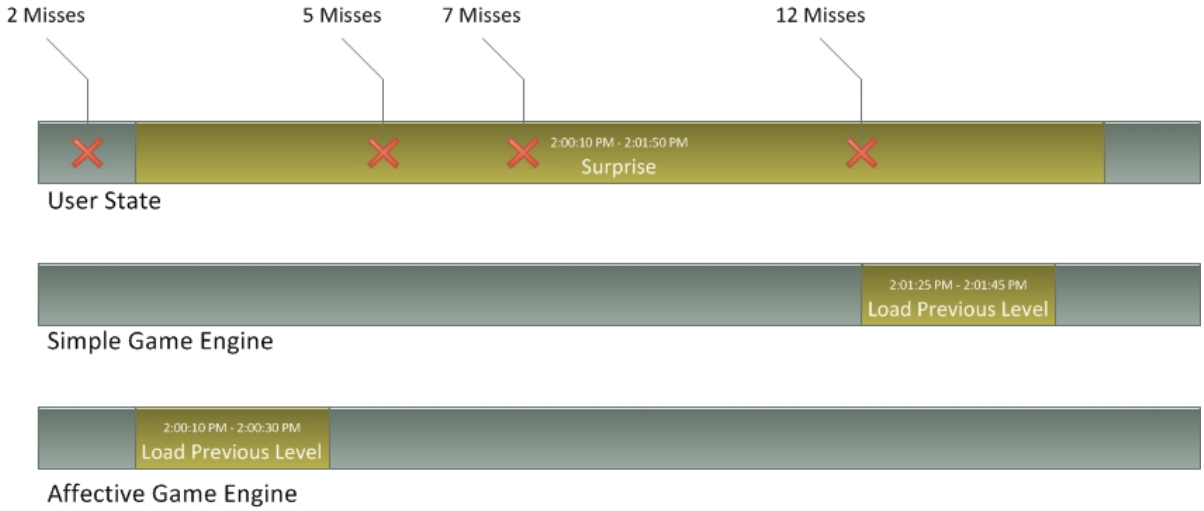


Figure 22: Difference between Simple Game Engine and Affect-mediated Game Engine. The Affect-mediated Game Engine will load the previous level first at the first sign of a sustained negative emotion. However, the Simple Game Engine will wait until the user has missed enough items.

Debug mode

For debugging purposes, we included the ability to display debugging information in two forms: a full debug string and a hidden emotion letter. The debug string, which can be displayed at any time during the game, will show the current score, the current level, and the user's current sustained emotion (Figure 23). The emotion letter is hidden away in the bottom left hand corner of the screen. It is always present, and was used primarily for the purpose of secretly assessing the emotional detection accuracy of the Game System during another user's play (Figure 24).

Figure 23: Debugging information that displayed (from left to right) the score, the current level, and the user's current sustained emotion.

Figure 24: A letter representing the user's current sustained emotional state. This letter was used during the study to quickly assess the game's detection accuracy while a participant played the game.

9 Experimenting with the Game System

Overview

We wanted to know whether our affect-sensitive game system actually helped its users move through the various levels to beat the game, and thus, we tested it through an experimental study. Two experimental groups were established: an affect condition and a control condition. In the affect condition, the Affect-mediated Game Engine is used (recall, this engine not only records the user's facial expression but also adjusts itself based on that expression). In the control condition, the Simple Game Engine is used. We believed that participants would perform better in the affect condition than in the control condition. Specifically, we expected that the scores would be higher and the game's play duration would be shorter in the affect condition than in the control condition.

We performed experiments at the Children's School. The Children's School is an educational institution part of Carnegie Mellon University that allows students and faculty the ability to conduct research and observational studies on child development in a controlled environment. The School admits children in three programs: a three year old preschool, a four year old preschool, and a kindergarten (five year olds). All children are eligible to take part in research studies, but can choose before or during a study to leave. All projects that take place in the Children's School must be approved by the school's director, Dr. Sharon Carver. Subsequently, research studies that fall under "human subject research" must attain IRB approval.

Table 10: Observation Results

Activity	Facial Expression
creating something (in the middle of task)	neutral (emotionless) expression
completion of task, convergence to some realizable product	happy expression (eureka moment)
high confidence on task	(sustained) happy expression
hesitation	change from happy expression to neutral expression
inability to keep up with task	worried expression

Pre-experiment observation

Before designing the research experiment, children in all age groups were observed in their respective programs to establish a range of facial expressions that they expressed during various activities. The results of this observation informed the experiment design by deciding which affective states our game would be able to detect and in which situations these affective states occur (Table 10). With this information, we attempted to design our game to adapt to these affective states and to explicitly evoke certain feelings within the children.

In our observation, we found that the children were happy, worried, or neutral in their activities. Happiness was characterized as either laughter or a sustained smile, and often came about through satisfaction from completing a task. A worried expression was very much similar to surprise, as the child's mouth would open wide in those moments in which there was hesitation or if he felt that he could not keep up with a task. For the most part, children showed no emotion (neutral) when they engaged in an activity. Furthermore, children transitioned between these expressions with their mastery of their activity. For example, a child might start out with no emotional expression or with a happy expression and move to a state of worry if he begins to struggle with an activity.

Pilots

After designing our game to this observed range of emotions, we conducted a few pilots to acquire feedback on the game's usability and difficulty (and to also detect any unintended bugs or glitches). Five year old participants were chosen by staff members to play our game (the selection was arbitrary, based mostly on if a child was not already occupied with an activity). The pilots resulted in improvements in the visual design of the game, the usability of the mouse as the primary mechanism of interaction, the game instructional script, and the game's difficulty.

Design revisions

Figure 25: Old design using ribbons to distinguish colored baskets. Active basket locations were random, sometimes sandwiching inactive baskets. Also, we used a vertical progress bar.

Our initial game design used colored ribbons to better distinguish the five colored baskets (Figure 25). While this design was elegant, our participants had difficulty making the connection between each colored basket and the associated colored item that fell. A lot of pilot participants used arbitrary baskets to catch food items. Furthermore, some

participants seemed very much distracted by some of the baskets that were deemphasized in view (not active in the current level) (Figure 26). It was thought that the random arrangement of the active baskets would sometimes sandwich the inactive baskets, further emphasizing what should be mostly hidden.

We also felt that children were ignoring our vertical progress bar (Figure 25), which might have contributed to the aimless gameplay. This was immediately fixed, and the horizontal progress bar was born Figure 26). We figured placing a progress bar at the bottom of the screen would be more obvious to the participant as a lot of the action (selecting a basket/items hitting the ground) happened towards the bottom.

Figure 26: Here, the active basket locations are seen sandwiching inactive baskets. The vertical progress bar is replaced with a horizontal bar at the bottom.

This modification further confused children who could not understand the icons we chose to represent the start and end of the progress bar (a frown face and a checkered flag). In particular, some questioned whether the game was a race since a checkered flag is usually associated with the end of a race.

In our final iteration (Figure 27), we placed the food items directly onto the basket so that the children could better make the association. We also made sure that the baskets

became active in order from left to right so that an inactive basket would never rest between two active ones. Lastly, we removed the start icon and replaced the checkered flag with a star. We felt the star in general represented something ”good” that the children would want to attain.

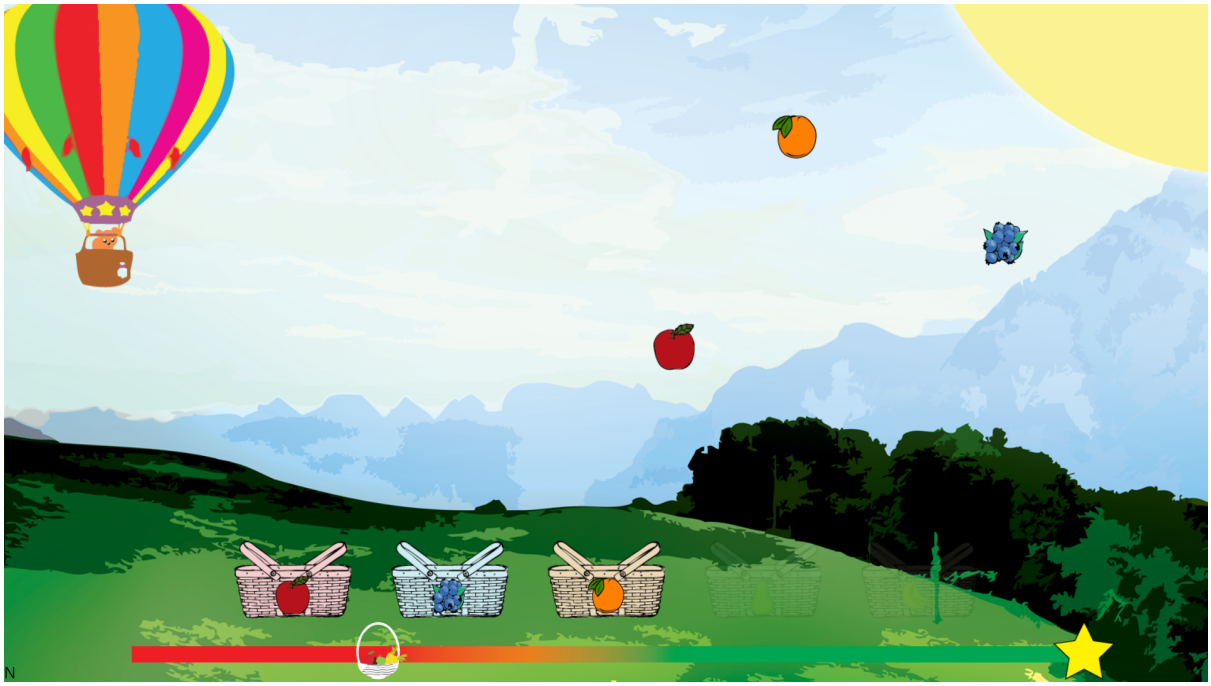


Figure 27: Final iteration. Pictures of food items are placed directly on baskets. Baskets appear in order from left to right. A star is used to designate the ”win” state.

Mouse improvements

We noticed early in our pilots that children were having difficulty interacting with the game using our mouse. First, they could not keep track of the mouse cursor on the screen as it would fly from side to side. It also blended too well with the game's background. Also, they would sometimes pull the mouse too far off the table such that its optical laser would stop tracking the surface, freezing the cursor in its place on screen.

To remedy these mouse issues, we increased the size of the cursor and changed it from white to black so that it would stand out more. We also slowed down the cursor to about half the original speed. Lastly, we wrapped the mouse cord around the laptop

to restrict wide movement of the mouse so as to deter children from pulling it off the table.

Better instructions

We initially told participants that they simply had to catch the food items in order to win. We had assumed that they would make the connection between catching/missing an item and the sliding of the basket in our progress bar. Unfortunately, it seemed like the participants were just randomly choosing to catch items, not knowing exactly when the game would end and how the game would end. Thus, we made sure in the actual study to very slowly and carefully explain every aspect of the game, specifically pointing out the progress bar and how each catch or miss effects it. We also emphasized the goal of the game - to make the basket in the progress bar "touch" the star. Please refer to Appendix 12 for the final version of the instructions.

Game play

The game was designed so that participants would have to catch 80 items in order to win. Participants who seemed skillful at playing the game seemed to take a very long time to finish it because of this requirement. We felt that the less skillfull players might get discouraged if they could not make progress in the game within a reasonable time period. Thus, we lowered the max items requirement down to 70 (and then subsequently down to 60 after gaining more info during the study).

We also adjusted the speed at which items fell in the various levels. The speed of the last level was decreased slightly since some participants reached that level but struggled to move past it and win the game — we wanted participants to be able to win. In the actual study, however, the speed of the levels was adjusted again such that the first two levels ran at a moderately slow speed, the middle two levels ran at a medium speed, and the last level ran at the initial fast speed.

Participants

The study utilized five year old and four year old children participants from the Children's School. We had intended on choosing subjects from the three year old age group, but decided, after witnessing our four year old participants struggle in the game's early levels, that they would have a much harder time playing our game. It might be worth in subsequent experiments to test this assumption, however.

Participants were selected based on their availability on testing days. Furthermore, we could only test on days when the testing room was free for use from other researchers.

Ultimately, eighteen boys and nine girls participated in the study, four of which were of asian ethnicity, two of african-american, and the rest of white-caucasian. Eleven of the participants were five year olds and the rest were four year old. In total, 27 subjects participated.

Setting and Procedure

Environment

Experiments were carried out in two locations at the Children's School. Because the dedicated experiment lab rooms were occupied during our study, we were allowed to conduct our experiments in the corner of the kindergarten classroom for the first round of five year old tests. For the remainder of the experiment tests, one of the lab rooms was used. The main difference between both experimental locations was the presence of other people: the classroom was much noisier and had several students running around doing activities while the lab rooms consisted of only the observer, researcher, and the participant. We did not feel the difference in location impacted a child's ability to play the game as the children remained focused in both settings.

Procedure

Each participant was given a basic overview of how to play the game where we focused on how to win and how points were gained and lost (depicted by our basket progress bar) (see Appendix 12). During this explanation, we demonstrated each aspect of the game in front of the participant. We then let the participant try to play for a while (from where we left off in the earlier levels) to get acquainted with the mouse and the game's mechanics. When we confirmed that the participant understood how to play, we restarted the game and let the participant play from the beginning.

We wanted some way of assessing the accuracy of the system's detection of the player's facial expression. In some sessions, an observer sat in front of the participant and made note of his own guesses at the participant's current emotional state. The observer was not trained in Ekman's basic facial expressions and so did not pick up many of the seven expressions, but rather made complicated assessments. For example, he would say that a participant was "determined" or "concentrating" at times. The observer tracked every change in emotional state so that we could compare his sequence of emotional states with the sequence captured by the system.

We had wanted to be able to compare how the same participant improved between the two conditions. Thus, participants played the game twice, and each play was in a different experimental condition. For example, some participants played the control condition first while others played the affect condition first. This measure was taken to remove any bias due to any particular condition played first. Both game plays also utilized a different theme. The first play was always the fruit theme and the second was always the vegetable theme. The difference in theme served to make the games appear different to the children who would be playing the same game twice and to remove any familiarity bias that would occur for playing in the same theme.

The amount of time spent in each experimental session varied. Due to Children's School research policies, our experiment had to be flexible to children wanting to leave in the middle because they either became bored or did not want to play anymore. Furthermore, we ended the experiments whenever we felt that a child was repeatedly

moving up and down the same levels and not making any progress. Thus, some children played the game for a shorter amount of time than others. On average, each experimental session took no longer than ten minutes, with the first two minutes dedicated to guiding the child through the game's concepts and the remaining for the child playing through the game on his own.

Side note

We had to tweak our difficulty settings earlier on in the study as we made new observations about our participants. While the pilot provided us with a sense of how well the children would perform in various levels of the game, we found that our first set of participants were all able to complete the game without any hesitation (they mostly lacked emotional expression, as well). We suspected that maybe our participants in our pilot were all low performers in terms of ability to play the game. Thus, we increased the speed of each level to make the game slightly harder for all potential participants. However, our next set of participants struggled greatly to get past the early levels of the game with this new adjustment. This event prompted a balance between the easier settings and these much harder settings, and the resulting moderate settings proved to work well in subsequent experimental sessions. Even if the difficulty of the game differed between our first set of participants and our remaining participants, we could still analyze the results as the data collected between experiment sessions of the same participant were sufficient to answer whether a child improved in the affective condition.

Results

Improvement can be judged by many factors. We chose to look at the absolute score, the catch/total ratio, and the struggle value.

Score

The max (peak) score was captured for each child (recall that the score is the total number of catches minus the total number of misses, above 0). Scores from the early testing sessions with different settings were normalized to fit within the score range of the rest of the sessions. A higher score is a clear indicator of high performance, and in some cases may designate a win. On average, we found that childrens' max scores in the affect condition (n = 23) were slightly higher than those in the control condition (n= 22): approximately 34 for affect and 31 for control (see Figure 28). Using a t-test assuming unequal variances, it was found that the increase was not at all statistically significant at the 5% level (p-value was 0.257), however.

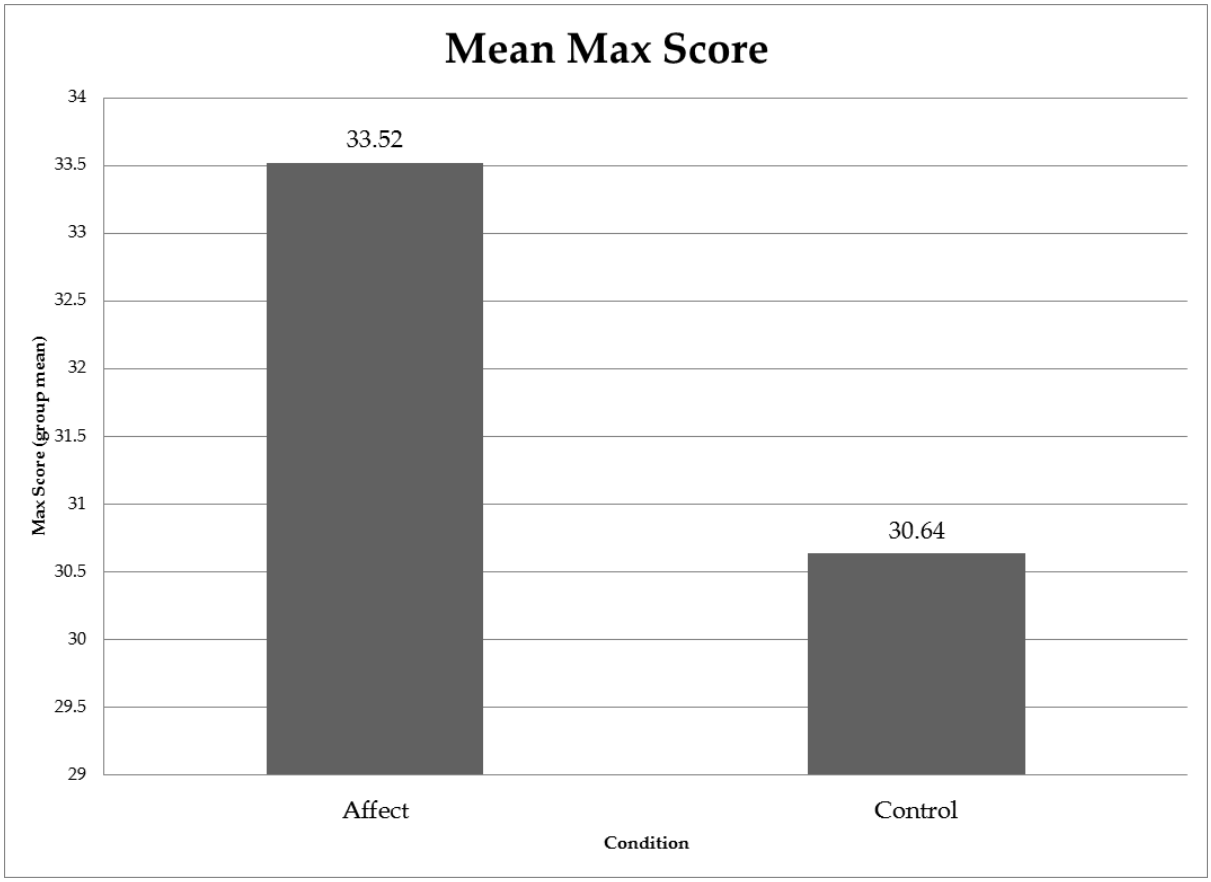


Figure 28: The average peak score across conditions.

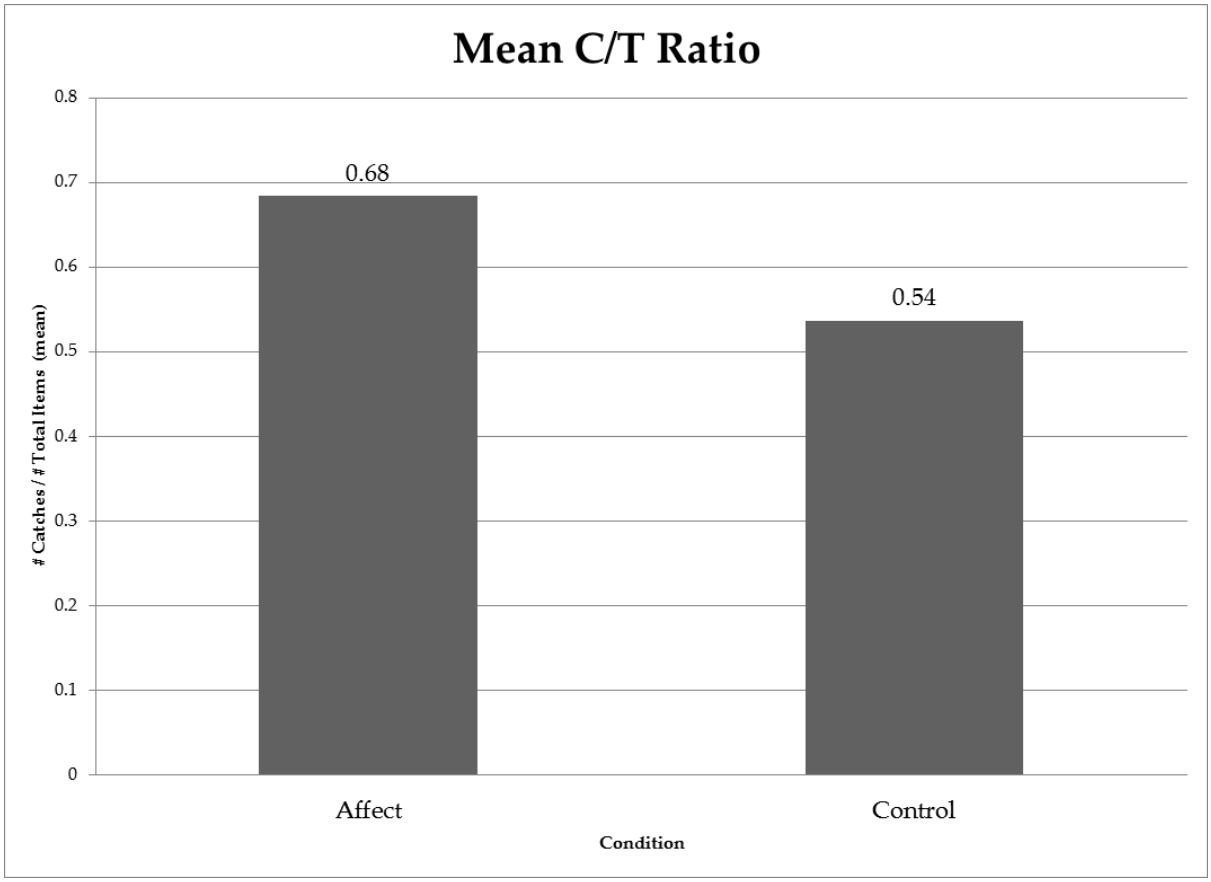


Figure 29: The averate catch/total ratio across conditions.

Catch/Total ratio

Another metric for measuring game performance is C/T Ratio, the number of total item catches over the total number of items spawned in the game session. A high C/T ratio indicates more items caught within the session, and thus, may represent higher game performance. We calculated the means of C/T ratios for individuals in each condition (see Figure 29) and compared them using a t-test assuming unequal variances . We found that the mean C/T ratio in the affect condition (n = 21) was higher than that in the control condition (n = 20) (approximately 0.68 to 0.54) and the results were significant at the 5% level (p-value 0.0059).

It is important to note that these results excluded three outlier data points (from both conditions) that had excessively long game session times. We had intended to not disturb a child that seemed engaged and determined to win the game, but it was later found that some children may politely continue to play a game regardless of how well

they perform and despite their actual desire. In each of these sessions, the child's performance became worse as the game progressed far past the average game time.

We also wanted to know whether the same subject improved across conditions. Using a paired t-test, we found that the same subject in the affect condition in general had a higher C/T ratio (by 0.13) than in the control condition (n = 17) (Figure 30). This result was statistically significant at the 5% level (p-value was 0.0011), suggesting that the subject did improve in the affect condition.

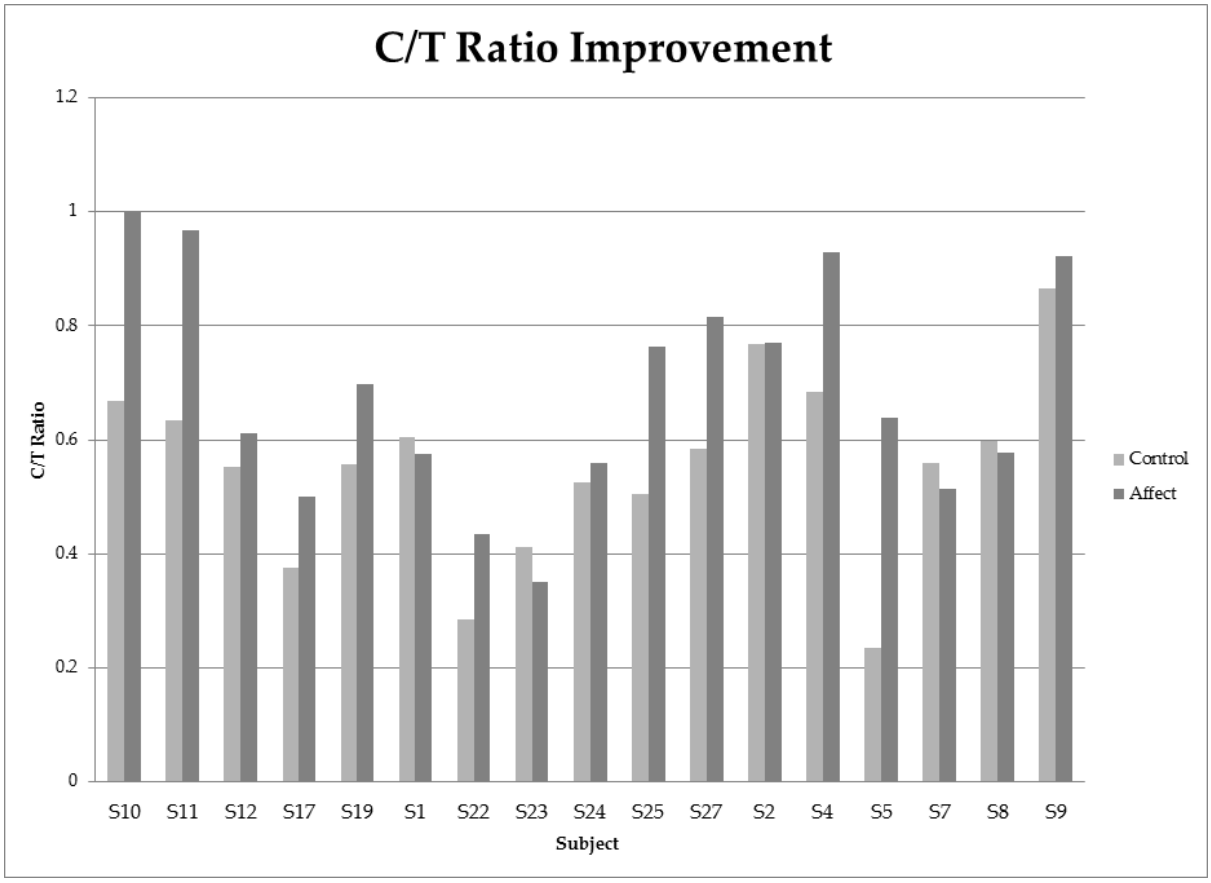


Figure 30: The catch/total ratio for the same subject across conditions.

Struggle

We captured the participants' difficulties in catching food items by measuring their struggle counts. We define a struggle as every moment a catch is followed by one or several misses. The total number of struggles within one session is not the same as the total number of misses as each struggle implies an ongoing attempt to catch items that

fall, whereas a set of misses alone may be caused by a participant either giving up or taking a break to survey the current game state.

Excluding the same three outliers from before and those data points from the earlier sessions with different settings, we calculated the mean total struggle across both conditions (figure 31) and found that on average, there was less total struggle in the affect condition (n = 15) than in the control condition (n = 18) (approximately 18 to 24). Furthermore, the results in each condition were compared using a t-test assuming unequal means at the 5% significant level and were shown to be almost statistically significant (p-value 0.0512).

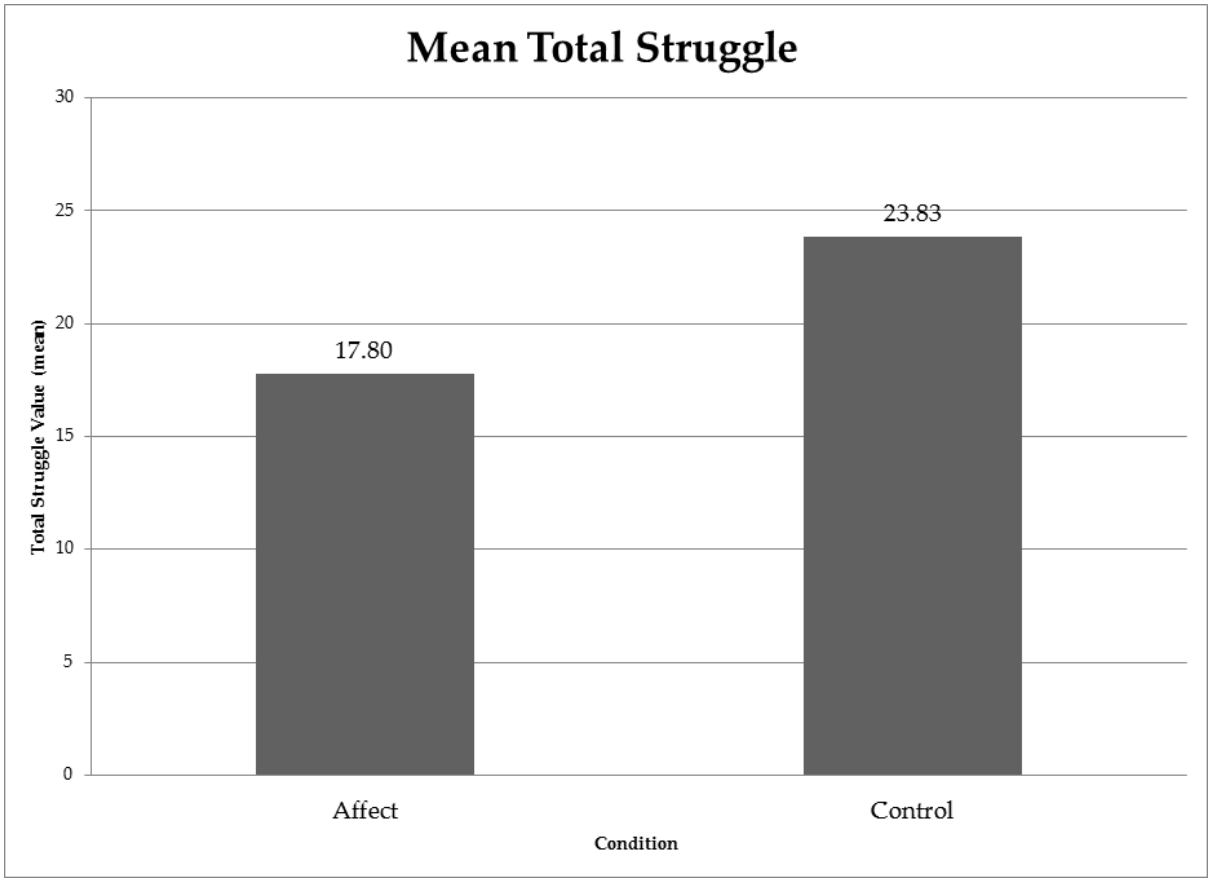


Figure 31: The average total struggle across conditions. A struggle is defined as a catch followed by one or several misses.

We separated the struggle counts by the level in which they occurred (see figure 32). At a first glance, it seems that there is more struggle in the higher levels in the control condition (n = 22) than in the affect condition (n = 23) (particularly in level 4). Similarly,

there appears to be more in lower levels in the affect condition. Only the difference in level 2 is significant according to the t-test (p-value 0.0131). However, at significance level 10%, level 4 and level 5 differences between the two conditions may be significant (p-values 0.0692 and 0.0648 respectively).

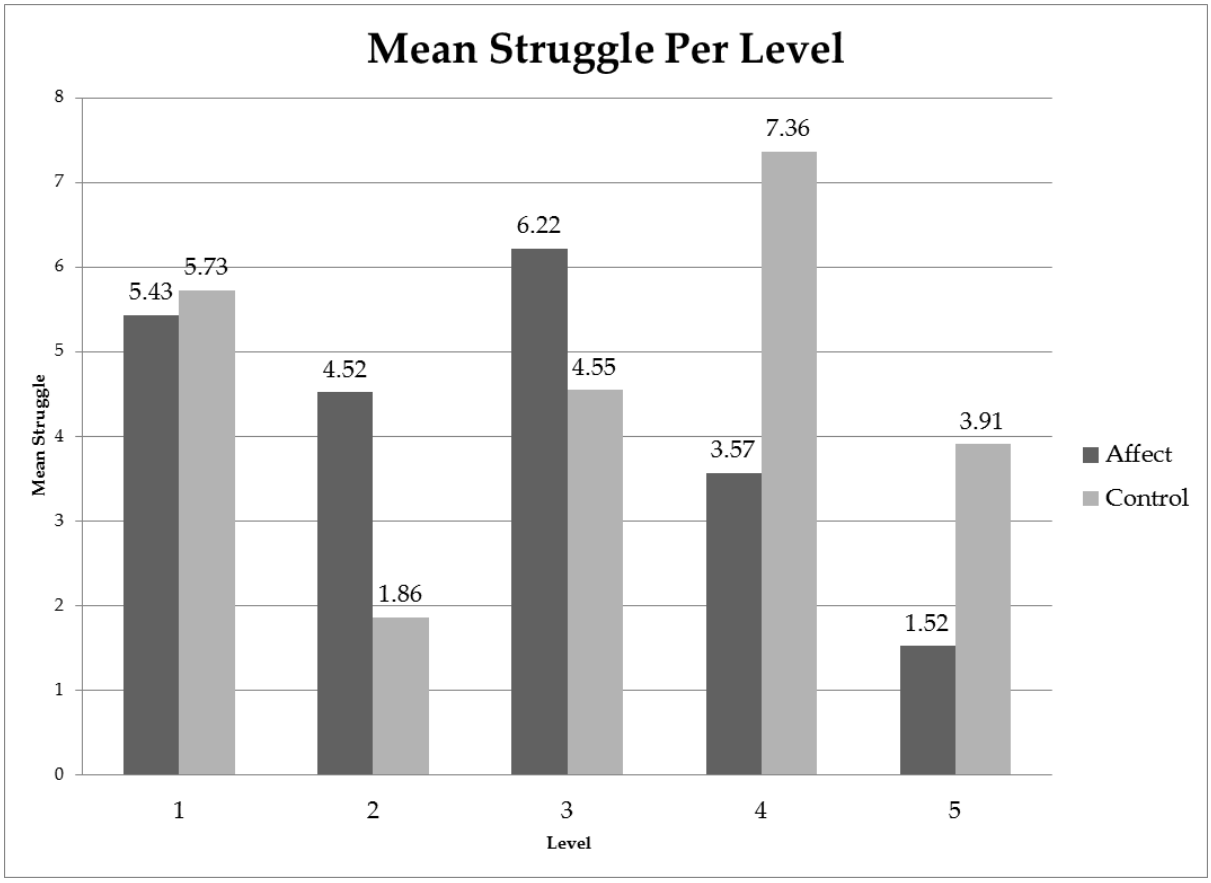


Figure 32: The average struggle broken down by level across conditions.

It also seemed as if the struggle counts were more spread across levels in the affective condition rather than placed in one or two levels. To test this hypothesis, we calculated the standard deviation of the counts across the levels for all data points. We then performed a t-test assuming unequal variances to compare the standard deviations between the affective (n = 21) and control (n = 22) groups (figure 33). The results of this t-test show some significance (p-value 0.00138) that the condition had on the spread of struggle value.

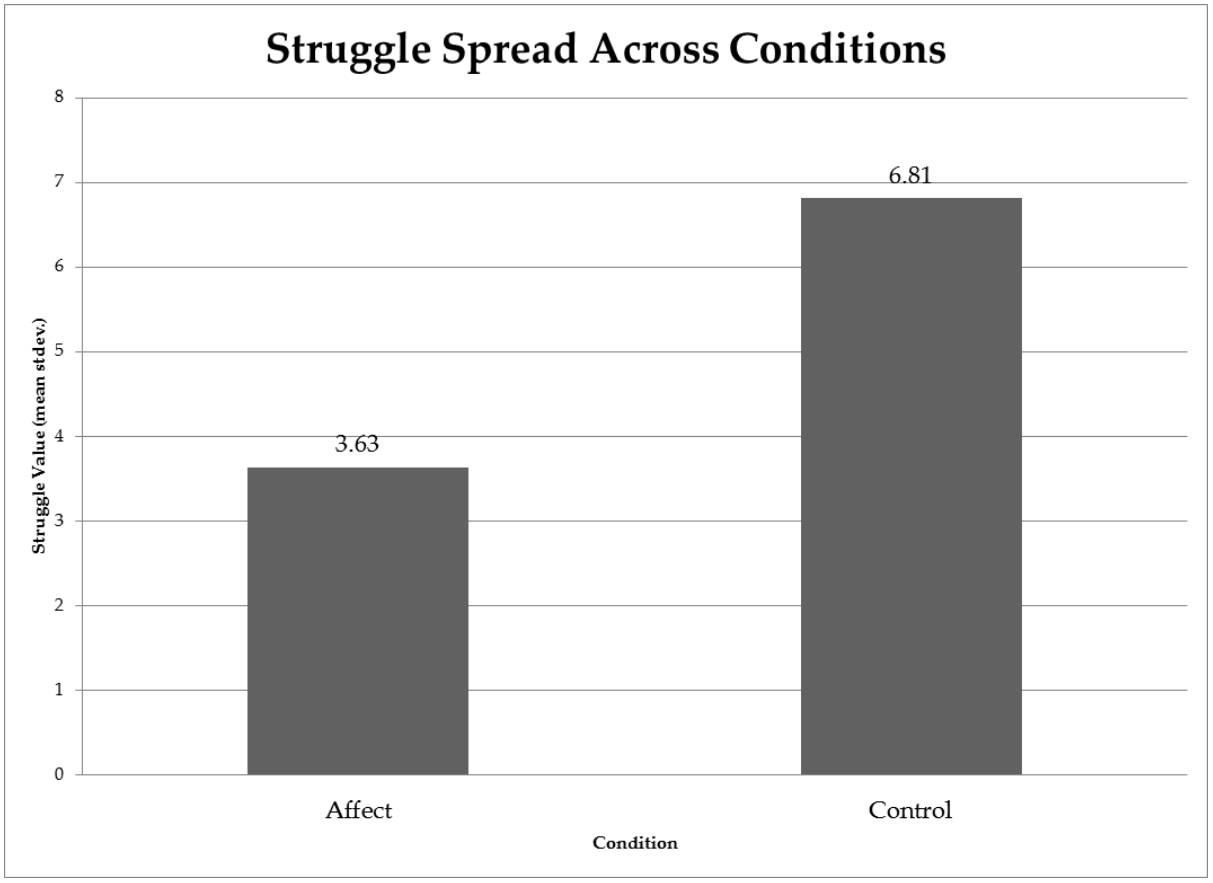


Figure 33: The average struggle spread (standard deviation) across conditions.

Recognition Engine Accuracy

We compared the notes made by our observer in some experimental sessions with the Facial Expression Recognition Engine's assessment (Table 11). For the most part, both were in agreement if we consider more complex states of "concentrating" and "determined" as still being represented as neutral expressions. However, it did seem like the Recognition Engine was confusing a lot of other emotions with neutral. In particular, we observed during a lot of experimental sessions that the Recognition Engine would guess that the child was showing disgust when he was actually neutral.

Improved Emotional State

We were also curious whether students were overall happier in the affect condition than in the control condition. We counted the number of instances of each emotion that

Table 11: Observer - detector agreement over user's emotional state.

Observer Notes	Detector in agreement?
Neutral, concentrating	Yes
Neutral, concentrating	Yes
Neutral	Yes
Neutral	Almost. Neutral was most common, but detected anger as well
Neutral, determined	Yes
Neutral (Smile)	Almost. Neutral was most common, but detected anger, disgust, and surprise as well
Neutral, determined, concentrating, smile	Almost. Neutral was most common, but detected anger as well
Neutral, concentrating	Almost. Neutral was most common, but detected surprise as well
Neutral, smile, frustration, neutral	Yes
Neutral, frustration, neutral, smile	Yes
Neutral	No. Detected neutral, anger, surprise, and disgust.
Neutral	Almost. Neutral was most common, but detected disgust as well.
Neutral, smile, frustration, smile, bored	Yes
Neutral	Almost. Neutral was most common, but detected disgust as well.
Neutral	Neutral, surprise, disgust
Neutral	Almost. Neutral was most common, but detected disgust as well.
Neutral, determined	Yes
Neutral, distress	Yes

were the most common and second most common per participant in each condition (Figure 34). Overall, sadness, fear, and contempt were not displayed. On the other hand, neutral (lack of emotion) and disgust were prevalent, followed by some anger. In the affect condition, happiness was found to be a second most commonly expressed emotion. However, this happiness was exhibited by only one person, and thus the result is not statistically significant. We also note that disgust was detected by less participants (6

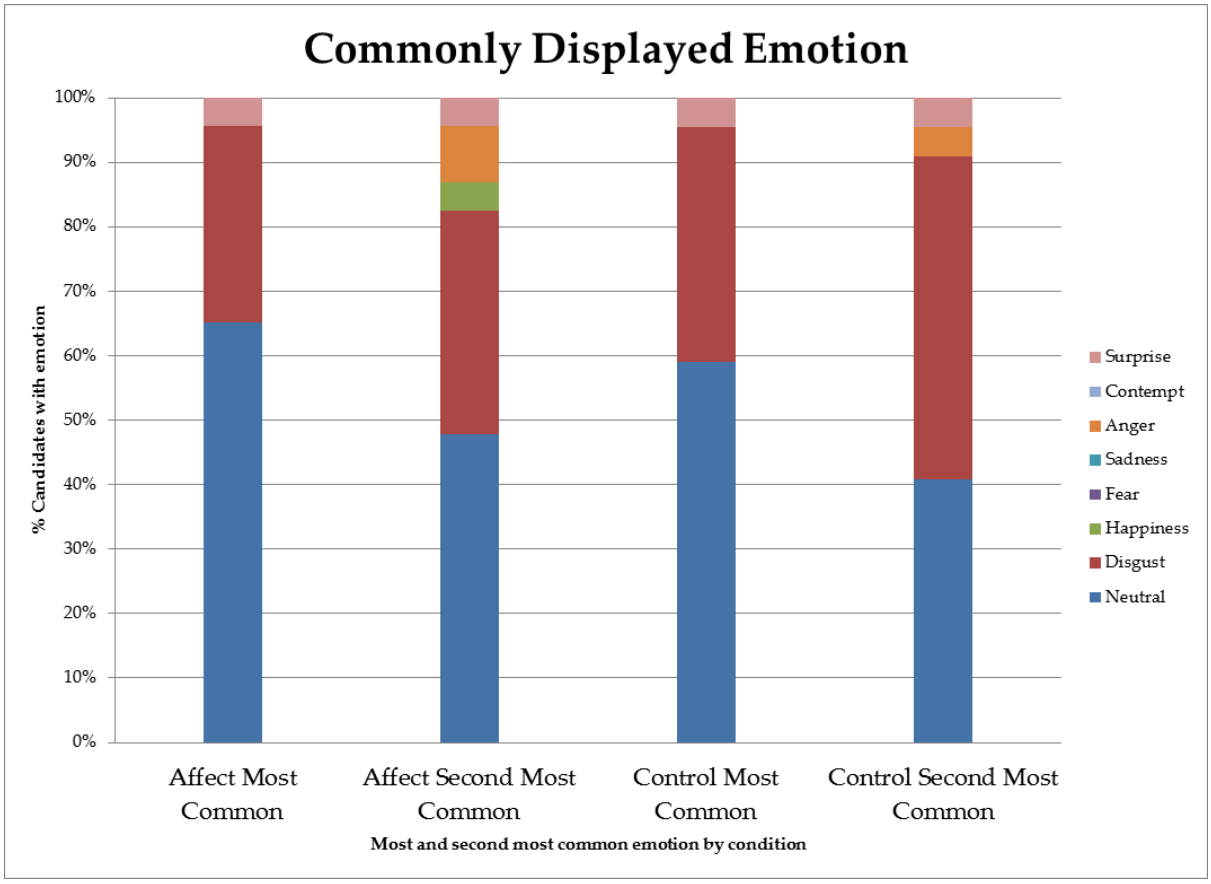


Figure 34: The commonly displayed emotion across conditions.

Other observations

It is interesting to note that only our five year old subjects actually won the game. The fastest five year old won in 2.4 minutes, and on average it took 2.9 minutes to win in the affect condition (n = 4) and 4.9 minutes in the control condition (n = 4). None of the four year olds actually won, which indicates a substantial difference in skill level

with just one year of age.

Discussion

There definitely does appear to be some improvement in performance in the affect condition. While the max score differences were not statistically significant to differentiate the two groups, the C/T ratio and the struggle spread do indicate that the affect-sensitive version of the game may better prepare a participant for victory.

In particular the C/T ratio was higher in the affect condition (especially for the same subject). The higher ratio could be the result of becoming more accustomed to the game. The affect-sensitive component of the game system aims to detect when the player is feeling overwhelmed or is in distress, and will move the game back one level upon any of these events. We would expect the system's behavior to reduce the player's stress so that he could regain composure to better grasp the current game state. Once the player has calmed, he may begin to catch more items than before in the harder level.

On the other hand, we do assume that a state of overwhelm or of distress is matched with a set of misses. One concern we had was that the affect game engine would simply load the previous level immediately before the player missed any items. Thus, a player could increase the number of catches, move on to a harder level, show distress, move back one level, and then catch more items, thereby always only catching and never missing. This scenario would then contribute to a much higher C/T ratio. However, we note that the game engine does not immediately respond to brief changes in facial expressions - the expressions must be **sustained** for some time before the system will designate the participant as expressing that emotion. Thus, a participant must show fear or surprise for a longer duration, and the only way for this to happen is through contstant interaction with the game's state - the inability to catch the various items that appear on the screen, and the subsequent misses of these items.

Flow

One interesting result from this study was the concept of struggle spread, more specifically that a participant's amount of struggle was more evenly spread across the game's levels and not piled up on one particular level. We felt that this spread could contribute to better engagement throughout the game. For example, one who is skillful might wait for a long time for the next item to fall because he has already swiftly collected all the items on the screen. In this time period, his mind might wander and he may detract away from the game. However, he might commit his mind to the game if there were reasonable bouts of struggle throughout.

We may go further and suggest that this active level of engagement is essentially a state of flow. Recall that a person is in flow when there is both high level of challenge and high level of skill (Csikszentmihalyi, 1990). We can consider a steady distribution of struggle throughout the game as indicative of a challenging activity. Furthermore, the participant must have sufficient skill if he can move between these levels. Thus, the participant who is sufficiently (but not overwhelmingly) challenged throughout the game might be in a state of flow.

Further experimentation would be needed to examine the effects of an affect-mediated system on a user's flow.

Insufficiency of Single Detectors

We noticed in some cases the system would confuse a child's neutral expression with one of disgust. It is possible that participants in the affect condition who won might have finished the game sooner if the game had not falsely detected an expression of disgust and moved the participant back one level. One explanation for this detection inaccuracy could be that the attached camera could not sufficiently construct a face model for the child's face, and thus, was contorting the face model to create an expression of disgust. If the child's face moved to the edges of the camera's view, the face model might warp to create this faulty face model fit. Specifically, we noticed that

some children rested their heads on their arms. If their hands blended too well into their face, the facial expression recognition engine might consider the hand as part of the face and will attempt to wrap the face model around the entire head-face component. Thus, the accuracy of the expression detection mechanism would depend on the stability of the face in the image.

It was also clear to us that the system might still miss many rich displays of emotional expression, specifically those involving the rest of the body. Our participants expressed their emotions through more body parts than just their faces. We saw children sighing and slouching down in their chairs. Some participants shouted, "Oh no!" while others gleefully stated that an item was their favorite fruit or vegetable. When a level became too intense for a few participants, those children leaned forward, ready to respond to the next falling item. In many of these cases, their faces showed no emotional expression, and thus the game system simply did not react accordingly. Thus, we wonder how much better the affective game engine would perform if it could capture the body as a whole.

Part V

Conclusion and Future Work

10 Study Summary

We explored the use of affect detecting mechanisms in a computer game system we designed and developed. The game, called the BasketGame, was designed for young children and required that players catch falling items of various colors with their associated baskets. The backend to the entire system used facial expression recognition (one mechanism for detecting emotion) to estimate a user's emotion to communicate back to the game. The system then altered the game based on this emotion, making it easier for the player when it detected emotions related to distress or frustration. The intent was for the system to better prepare the player for the much harder levels of the game so that he could win.

We conducted a study using this game with children at the Children's School and found favorable results across two conditions: the affect condition, in which the game will change itself based on the player's emotion, and the control condition. On the basis of simple catch performance, children in the affect condition tended to have a higher catch ratio (Figure 29, Figure 30) (determined by dividing the child's total number of food item catches by the total number of items dropped) than those in the control condition. We also assessed the performance based on the participant's struggles (a struggle is a catch followed by one or many misses). There seemed to be less struggles in the affect condition than in the control; moreover, these struggles were more spread out in the affect condition (Figure 31; Figure 33). We concluded that the affect-sensitive components of the game system did help participants adjust to the game's difficulty and could potentially enable flow (Csikszentmihalyi, 1990).

Our results also suggest that facial expression recognition alone is not sufficient to provide a guess of the user's emotion. While it can reliably classify a user's facial expression, the majority of the face still needs to be present to get a decent reading. If some parts are occluded or not in the camera's view (for example, some children rested their heads on their hands or slouched down past the edge of the camera), the facial expression engine may attempt to fill in those gaps, which may result in an inconsistent facial expression. Furthermore, if the entire face is not in the camera's

range, then no expression detection can be performed at all.

We also note that emotion can be expressed in many different ways and that our study only looked at one source, facial expressions. A person's posture, whether he is leaning forward or has his hand on the side of his head may indicate anxiety, and the higher pitch in his voice might represent joy. During the study, we felt that some participants, while maintaining neutral facial expressions, expressed distress through the rest of their body. The system that could only observe the user's face missed these subtle emotion markers all over the body.

Thus, more detection mechanisms are needed to generate a more comprehensive view of the user's emotional state that any single detector might lack.

11 Future Work

A follow up to this study should seek to expand the technological underpinnings of the system and the experimental power. First, the system should incorporate more detection mechanisms beyond facial expression. For example, we may augment the system with the body posture pressure measurement system (BPMS) that AutoTutor employed (D'Mello et al., 2008) to better detect the user's shifting posture within his seat. For example, we could use a body pressure to capture those moments when a child's level of engagement increases. Furthermore, we may utilize Microsoft's Kinect to dissect the user's full skeleton to recognize particular gestures or other physical descriptors, such as when one places his head on his hand because he is bored (Shotton et al., 2011). Other improvements can be made specifically to the main camera. For example, a more mobile camera that can turn to track the user might increase detection accuracy as the user will continuously remain in sight no matter how he moved (if he slouched down past the camera's initial view, for example).

Second, a much larger sample set may garner more specific results. This study was only exploratory, as it served to demonstrate how an affect-mediated system can help a user complete tasks. Thus, the results utilized a smaller and more predictable sam-

ple set: we used children as our participants because we had assumed that they would not regulate their emotional expression and would more obviously convey their emotions through large facial expressions. Thus, we should expand the participant pool to other subjects (especially after improving the detectors). We would also want to examine specifically where it might be appropriate to augment a machine with affect-sensing capabilities, and in particular, what we should expect from such a machine (for example, under which conditions would our affect-mediated system be able to sustain flow?).

Part VI

Appendix

12 Sample Game Introduction Script

Fruit Theme

This is Billy. Billy's fruit are falling from the sky. Your job is to catch the fruit like this (demonstrate by catching a few fruit of different colors). See, everytime you catch a fruit, this basket at the bottom (point to the progress bar) moves closer to that star (point to the star in the progress bar). When the basket touches the star, you win. But if you miss a fruit (wait for a fruit to hit the ground), it goes splat, like this (watch the fruit go splat) and the basket at the bottom moves away from the star (show the basket move away from the star when fruits hit the ground). So, you have to make sure you keep catching fruits. You try.

Vegetable Theme

This is Mr. Penguin. Mr. Penguin's vegetables are falling from the sky. Your job is to catch the vegetables like this (demonstrate by catching a few vegetables of different colors). See, everytime you catch a vegetable, this basket at the bottom (point to the progress bar) moves closer to that star (point to the star in the progress bar). When the basket touches the star, you win. But if you miss a vegetable (wait for a vegetable to hit the ground), it goes splat, like this (watch the vegetable go splat) and the basket at the bottom moves away from the star (show the basket move away from the star when vegetables hit the ground). So, you have to make sure you keep catching vegetables. You try.

13 Sample System Output

Table 12: Sample data output of the affect-mediated game system

Timestamp	Emotion	Level	Score	Pos. Streak	Neg. Streak	Total Catches	Total Misses
09:44:39	Disgust	2	13	5	0	16	3
09:44:40	Disgust	2	13	5	0	16	3
09:44:41	Disgust	2	14	6	0	17	3
09:44:42	Disgust	2	14	6	0	17	3
09:44:43	Disgust	2	14	6	0	17	3
09:44:44	Disgust	2	14	6	0	17	3
09:44:45	Disgust	2	15	7	0	18	3
09:44:46	Disgust	2	15	7	0	18	3
09:44:47	Disgust	2	15	7	0	18	3
09:44:48	Disgust	2	16	8	0	19	3
09:44:49	Disgust	3	15	0	1	19	4
09:44:50	Disgust	3	15	0	1	19	4
09:44:51	Neutral	3	15	0	1	19	4
09:44:52	Neutral	3	17	2	0	21	4
09:44:53	Neutral	3	18	3	0	22	4
09:44:54	Neutral	3	19	4	0	23	4

References

Chang, C.-C., & Lin, C.-J. (2011). LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, *2*, 27:1–27:27. (Software available at `http://www.csie.ntu.edu.tw/ cjlin/libsvm`)

Cootes, T. F., Edwards, G. J., & Taylor, C. J. (2001). Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *23*(6).

Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*. Harper & Row, New York.

D'Mello, S., Jackscon, T., Craig, S., Morgan, B., Chipman, P., White, H., ... Graesser, A. (2008). *Autotutor detects and responds to learners affect and cognitive states.* Montreal, Canada.

Ekman, P. (1972). Universal and cultural differences in facial expression of emotion. In J. Cole (Ed.), *Nebraska symposium on motivation* (p. 207-283). University of Nebraska Press, Lincoln.

Ekman, P., Friesen, W., & Hager, J. (2002). The facial action coding system. In *Research nexus ebook.*

Ekman, P., & Friesen, W. V. (1969). The repertoire of nonverbal behavior. *Science*, *164*(3875), 86-88.

Ekman, P., & Friesen, W. V. (1978). *Facial action coding system: Investigator's guide*. Consulting Psychologists Press, Palo Alto.

Ekman, P., Friesen, W. V., & Sorenson, E. R. (1969). Pan-cultural elements in facial displays of emotion. *Science*, *164*(3875), 86-88.

Ekman, P., & Heider, K. G. (1988). The universality of a contempt expression: A replication. *Motivation and Emotion*, *12*(3), 303-308.

Fox, E. (2008). *Emotion science*. Palgrave Macmillan, New York.

Gasper, K., & Isbell, L. M. (2007). Feeling, searching, and preparing: How affective states alter information seeking. In K. D. Vohs, R. F. Baumeister, & G. Loewenstein (Eds.), *Do emotions help or hurt decision making?* (p. 93-116). Russell Sage Foundation, New York.

Gross, R., Matthews, I., Cohn, J., Kanade, T., & Baker, S. (2008). *Multi-pie.*

Gross, R., Matthews, I., Cohn, J., Kanade, T., & Baker, S. (2009). Multi-pie. *Image and Vision Computing*.

Hansen, C., & Hansen, R. (1988). Finding the face in the crowd: An anger superiority effect. *Journal of Personality and Social Psychology*, *54*, 917-924.

Kanade, T., Cohn, J., & Tian, Y. (2000). *Comprehensive database for facial expression analysis.* Grenoble, France.

Lucey, P., Cohn, J., Kanade, T., Saragih, J., Ambadar, Z., & Matthews, I. (2010). *The extended cohn-kanade dataset (ck+): A complete expression dataset for action unit and emotion-specified expression.* San Francisco, USA.

Neisser, U., & Harsch, N. (1992). Phantom flashbulbs: False recollections of hearing news about the challenger. In E. Winograd & U. Neisser (Eds.), *Affect and accuracy in recall: Studies of f̈lashbulbm̈emories* (p. 9-31). Cambridge University Press, London.

Picard, R. R. (1997). *Affective computing*. MIT Press, Massachusetts.

Saragih, J., S.Lucey, & Cohn, J. (September, 2009). Face alignment through subspace constrained mean-shifts. *International Journal of Computer Vision (ICCV)*.

Sauter, D. A., Eisner, F., Ekman, P., & Scott, S. K. (2009). *Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations* (Vol. 107) (No. 6).

Shotton, J., Fitzgibbon, A., Cook, M., Sharp, T., Finocchio, M., Moore, R., ... Blake, A. (2011). Real-time human pose recognition in parts from single depth images. *CVPR*.

Vuilleumier, P., Armony, J., Driver, J., & Dolan, R. (2001). Effects of attention and emotion on face processing in the human brain: An event-related fmri study. *Neuron*, *30*, 829-841.

Winkielman, P., & Trujillo, J. L. (2007). Emotional influence on decision and behavior: Stimuli, states, and subjectivity. In K. D. Vohs, R. F. Baumeister, & G. Loewenstein (Eds.), *Do emotions help or hurt decision making?* (p. 69-91). Russell Sage Foundation, New York.