Proceedings of the Twenty-Seventh International Florida Artificial Intelligence Research Society Conference

Using Affect Awareness to Modulate Task Experience: A Study Amongst Pre-Elementary School Kids

Vivek Pai and **Raja Sooriamurthi**
Carnegie Mellon University
5000 Forbes Avenue,
Pittsburgh, PA 15213

Abstract

Computing systems may record facts about users such as their click behaviors, their accuracies, and their rates of recurrence, and may then act on predictions of their behaviors and preferences. However, sequences of key presses and mouse movements do not capture their immediate desires and needs during a task. Affect-mediated systems, computing systems that adapt to the emotional state of users, can better respond to these momentary shifts in demand. This paper presents a study of how affect mediation can help improve a user's task performance. Designed as a children's game, our experimental system uses facial expressions to regress the user to an earlier, easier game phase according to perceived unease. Through experimentation with child participants, we found that children performed better with the affect-sensitive version of the game than with the non-affect-sensitive version. We hope these results will support the future design of affect-sensitive machines that can help users complete tasks.

Figure 1: A screenshot of the affect-mediated game system, BasketGame. The goal of the game is to catch falling food items with their corresponding baskets. There are five different phases of the game with progressive difficulty. As the game proceeds, the current phase changes based on the player's facial expression.

Introduction

Artificially intelligent software systems are still limited in their ability to understand users and their motives while completing tasks. Artificial intelligence methods construct models from training data and observations and choose an action for a particular stimulus. These models may evolve based on direct feedback from the user and some measure of success. For example, an airline reservation system may use your travel history to suggest specific flights. If you choose not to select any of the suggested flights, the system will take note of this behavior so that it might adjust its future flight suggestions. The reservation system generates a model for the user for suggesting flights and adjusts it based on the user's response. But these models capture only the results of keystrokes and mouse clicks they do not express instances of a human's behavior expressed in-between or simultaneously with her input into the machine. More so, a model may "fit" the user's initially captured state, but the user's wavering mind may deviate from this initial state, leaving the model inconsistent and stale. The user might enjoy morning flights, but in this particular instance, she is much too fatigued to wake up early to catch this flight and might want a much later flight instead. Even if the artificially intelligent agent made a correct assumption of the user's behavior in that earlier state, 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.

We feel that analyzing the user's emotional state, in addition to her physical inputs, can fill-in these gaps in understanding the user's current intentions and desires. This paper presents an instance of what is known as an affect-mediated computing system, a system that can sense and respond to human emotion. The system, designed as a children's computer game, actively adjusts its model according to the facial expression a child displays. As the child struggles with the gameplay, as determined by expressions of disgust and surprise, the game will move the child to an easier phase so that she could prepare for the more difficult level. It attempts to help a child complete the game, which involves successfully passing through all phases of increasing complexity. We hope this system exemplifies how machines can better adapt to users during tasks by understanding their emotional states.

Emotion Recognition

BasketGame uses a facial expression recognition engine to adapt to its player's emotions. The engine is a modified version of Jason Saragih's FaceTracker, which performs face alignment and tracking using active appearance models (AAMs) that were pre-trained with images from CMUs MultiPIE face database (Saragih, S.Lucey, and Cohn September 2009; Gross et al. 2008; 2009). 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. LIBSVM was used to train the SVM model with a set of images from the Extended Cohn-Kanade (CK+) dataset (Lucey et al. 2010; Chang and Lin 2011). The training process used an RBF kernel and a 5-fold cross validation step, resulting in a model with a prediction accuracy of 70.59% in LIBSVM's testing. It worked well in informal tests to predict expressions of happiness, surprise, disgust, and anger. Lastly, SVM prediction code from LIBSVM was embedded directly into FaceTracker to allow for classification of facial model parameters generated by the original FaceTracker's alignment code. This modified FaceTracker became the engine we would use in our game system (Figure 2).

System Design

BasketGame was designed for young children and tailored specifically for the participants of our experiments. The goal of the game was to catch different colored food items (fruits) that fell with matching baskets (Figure 1). As the game progressed, the subtasks of catching these items became more complex as the rate at which the fruits fell, the number of different locations in which they fell, and the number of different items increased.

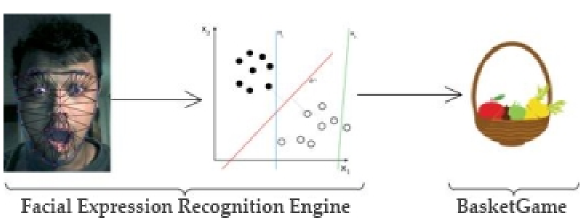

Figure 2: A high level overview of the system. The Facial Expression Recognition Engine processes video frames from a web camera, tracks the face in each frame, and classifies the face as expressing one of the seven basic emotions (or the neutral expression). It then sends this emotion label to the BasketGame to be used for manipulating the game's logic.

Gameplay

Specifically, players had to catch the food falling from the top of the game window with matching colored baskets resting at the bottom. They progressed through five phases of the game, which each described certain parameters: 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. Based on the player's accuracy in catching the food in the correct baskets in a phase, the game would move the player onto the next phase in which a greater variety of food would drop in more random locations, possibly at an increased rate. If a player performed poorly in a phase, however, the game would move the player back to the previous, easier phase. In order to win, players had to master all phases of the game, repeating those phases with which they struggled.

During experimentation, two different game engines were used: the Simple Game Engine (control) and the Affect-mediated Game Engine (experimental), the affect-sensitive component of the Game System. The game's gameplay was identical with either engine; however, how it moved the user through phases with each engine differed.

Simple Game Engine (control) With the Simple Game Engine, the game moved the player between phases solely based on the player's performance, which was determined by the player's consistent set of catches or misses. If the ratio of "catches to misses" within a level met a certain threshold, the game moved the player to the next level. However, if the ratio of "misses to catches" reached its threshold, the game moved the player to the previous, easier phase. These ratios were reset every time the game moved the player into a phase.

Affect-mediated Game Engine (experimental) The Affect-mediated Game Engine extended the Simple Game Engine by additionally reacting to negative emotional expressions. These negative emotions were determined before designing the experiment by observing the child participants' facial expressions during activities in their respective programs. We found that they were happy (through satisfaction in completing a task), worried (hesitating or frustrated within a task), or neutral (having steady engagement with the task) in their activities. Happiness was characterized through either laughter or a sustained smile, while a worried expression appeared through wide open mouths and eyes (almost resembling surprise). Furthermore, children transitioned between these expressions with their mastery of their activity. For example, a child might start out with a happy expression and move to a state of worry upon struggle with an activity. These expressions were then matched with the seven basic facial expressions and neutral expression to generate two sets: positive (happiness, content, neutral) and negative (fear, disgust, surprise, anger, sadness).

Thus, in terms of the Affect-mediated Game Engine, whenever a player exhibited an emotion from the negative set, the game would *immediately* move the player into the previous game phase, regardless of the player's catch/miss ratios. The game did not adapt to or accommodate positive emotions; in the case that a player exhibited an emotion from the positive set, the game would follow the default logic from the Simple Game Engine, in which the game moved the player between phases depending on the players "misses to catches" ratio.

The difference between the two game engines can be seen in figure 3. The user misses a few items and begins to show surprise. With the Simple Game Engine, the game will move

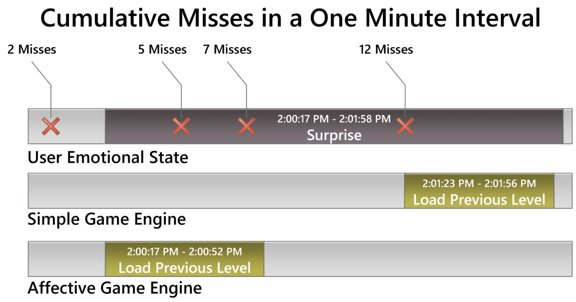

Figure 3: Difference between Simple Game Engine and Affect-mediated Game Engine. The Affect-mediated Game Engine will immediately move the player to the previous phase at the first sign of a sustained negative emotion. However, the Simple Game Engine will wait until the user has missed enough items.

the player to the previous phase only after the user misses twelve items, which satisfies the threshold for "misses to catches". However, the Affect-mediated Game Engine will move the player to the previous phase at the first detection of sustained surprise.

Winning A player won when the score, another running count, reached a max score threshold (which was 50 in our experiment), which was represented as a progress bar at the bottom of the screen.

Experiment

Two experimental groups were established: an affect condition and a control condition. In the affect condition, the Affect-mediated Game Engine was used (recall, this engine not only records the user's facial expression but also adjusts the game's state based on that expression). In the control condition, the Simple Game Engine was used (the engine that only responds to the user's performance). Our hypothesis was 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.

Setting

Experiments were performed at the Children's School, an educational institution part of Carnegie Mellon University that allows students and faculty the ability to conduct research and observational studies on child development with three, four, and five year olds in a controlled environment.

Participants

The study utilized five year old and four year old children participants from the Children's School. Children were selected within each group based on their availability on testing days.

Procedure

Verbal instructions Each participant was given a basic overview of how to play the game, specifically how to win and how the progress bar changed as points were gained and lost. Each aspect of the game was demonstrated in front of the participant. Then, the participant tried to play the early phase of the game to become 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.

Expression recognition validation 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 more general 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.

Participant improvement We had wanted to be able to compare how the same participant improved between the two conditions. Thus, participants played the game twice (across different days) in different experimental conditions, each time with either a fruit or a vegetable theme (to remove familarity bias and to make the game appear different to those children who might not want to play the same game twice).

Early termination 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.

Results

Improvement can be judged by many factors. We chose to look at the absolute score, the catch/total ratio, and the struggle count (Table 1).

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, in the range 0 to 50, where 50 is the winning score). 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

Measure	Description	Example	Interpretation
Score	total # catches - total # misses	40	High performance.
C/T Ratio	total # catches / total # items spawned	$\frac{\text{75 catches}}{\text{110 items spawned}}$	Lower relative efficiency.
Struggle	a catch followed by one or several misses	C,M,M,C,M,C	Some difficulty/unease.

Table 1: Descriptions of key result measures and some examples with their possible interpretations. A score of 40 (out of 50) could indicate above average performance. However, a C/T ratio of 75/110 means the player spent a longer time (than another player) to achieve that score, indicating lower efficiency. Lastly, two struggles (two catches each followed by a miss or two) shows she had difficulty keeping up in a certain phase.

(n = 23) were slightly higher than those in the control condition (n= 22): approximately 34 for affect and 31 for control. Using a t-test assuming unequal variances, it was found that the increase was not statistically significant at the 5% level (p-value was 0.257), however.

Catch/Total ratio

Another metric for measuring game performance is the C/T Ratio, the number of item catches over the total number of items spawned in the game session. A high C/T ratio corresponds to a greater number of items caught (and not missed) within the session, and thus, may represent higher game performance. We calculated the means of C/T ratios for individuals in each condition and compared them using a t-test assuming unequal variances . We found that the average 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 who 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 worsened as the game progressed far past the average game time.

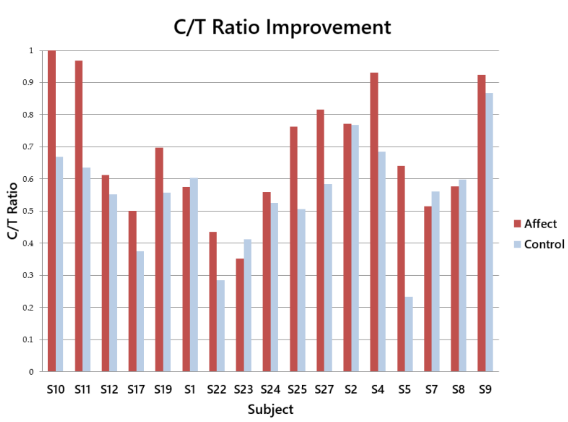

Figure 4: The catch/total ratio for the same subject across conditions. C/T ratios were found to be higher in the affect than in the control condition for the same subject.

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 (an increase by approximately 0.13) than in the control condition (n = 17) (figure 4). This result was statistically significant at the 5% level (p-value was 0.0011), suggesting that the subject did improve in the affect condition.

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 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).

We separated the struggle counts by the level in which they occurred (see figure 5). 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).

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 6). The results of this t-test show some significance (p-value 0.00138) that the condition had on the spread of struggle value.

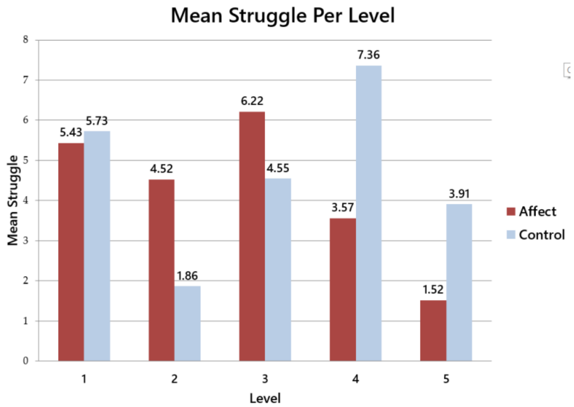

Figure 5: The average struggle broken down by phase across conditions. Participants struggled less in the harder phases (4 and 5) in the affect condition than in the control, but struggled more in the first 3 easy phases.

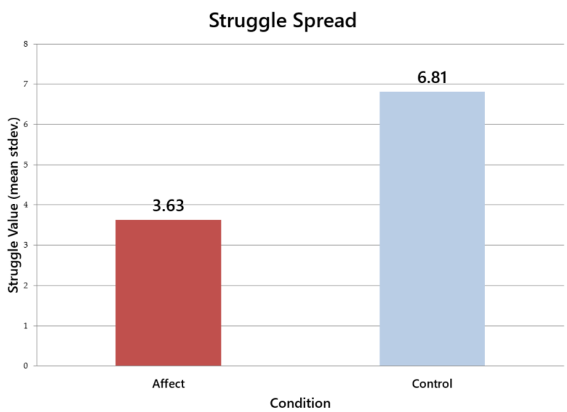

Figure 6: The average struggle spread (standard deviation) across conditions. On average, the struggle across the different phases was more spread out in the affect condition, possibly suggesting steady engagement in a slightly challenging task.

Recognition Engine Accuracy

We compared the notes made by our observer in some experimental sessions with the Facial Expression Recognition Engine's assessment. 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 dominated in each participant's game sessions. Overall, sadness, fear, and contempt were not displayed. On the other hand, neutral (lack of emotion) and disgust were frequent, followed by some anger. In the affect condition, happiness was found to be a second dominant 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% less) as a most common emotion in the affect condition than in the control condition. Unfortunately, due to sample size there were no statistically significant differences in emotion displayed between either condition.

More details regarding these results and a discussion of related research can be found in the original undergraduate thesis (Pai 2012).

Discussion

Children did appear to perform better in the affect condition. While the max score differences were not statistically significant between the two groups, the C/T ratio and the struggle spread indicated that the affect-sensitive version of the game may have better guided a participant through 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 aimed to detect when players felt overwhelmed or distressed, and would move the player back to the previous phase upon any of these events. We would have expected the system's behavior to reduce their stress so that they could more calmly progress through the harder phases.

On the other hand, we assumed that a state of overwhelm or of distress was matched with a set of misses. We were concerned that the affect game engine might simply move the player to the previous phase before the player missed any items. In this instance, a player could increase the number of catches, move on to a harder phase, show distress, move back one phase, and then catch more items in the easier phase, to always only catch and never miss. This scenario would then contribute to a much higher C/T ratio. However, this phenomenon was less likely as the game engine was designed to ignore brief changes in facial expressions - the expressions must be sustained for some time before the system designated 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 that a participant's amount of struggle was more evenly spread across the game's phases and not piled up on one particular phase. 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 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. According to psychologist, Mihaly Csikszentmihalyi, a person is in flow when there exists both a high level of challenge and a high level of skill during a task (Csikszentmihalyi 1990). Flow contributes to a person's productivity in task completion; one who is highly skilled and presented with less challenge might bore easily. Conversely, the less skilled worker in a highly challenging task might become demotivated. However, highly skilled workers are engaged in those tasks that are highly challenging, as if in a constant state of flow.

For BasketGame, flow might be realized through both the participant's struggle and performance. The steady distribution of struggle throughout the game can indicate a challenging activity, while the participant must show sufficient skill to be able to 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.

Conclusion

This study sought to explore how an affect detecting computer game system can modulate itself according to its player's emotional state. We attempted to pair a children's computer game with a facial recognition engine to provide the game with emotion sensing capabilities. Our study with this game system yielded favorable results. Players in the affective condition (when the game reacted to changes in emotional state) tended to perform higher than those in the control (non-affect sensing) condition. Affect participants also tended to struggle less; moreover, this struggle was more spread out than in the control condition (Figure 6). We concluded that the affect-sensitive components of the game system did help participants adjust to the game's difficulty, and its ability to spread the struggle throughout the game could potentially enable flow (Csikszentmihalyi 1990).

Future Work

A follow up to this study should seek to expand the technological underpinnings of the system and the experimental power. First, we would want to incorporate more detection mechanisms beyond facial expression. One thought could be to utilize Microsoft's Kinect to dissect the user's full skeleton to recognize particular gestures or other physical descriptors, such as when one places her head on her hand because she 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 a user might more accurately detect the facial expressions of the child who might have slouched down past the camera's initial view. Additionally, a larger, more diverse sample set may garner more specific results. For example, such a set might allow us 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?).

While our study focused on a single instance of aiding task completion, we see great potential in the future for machines that can adapt to their users' emotional states to help them complete tasks. One example of such a machine is an affect-sensitive GPS. You might hop into your car after a contentious project meeting to travel to your next destination. As you drive, your GPS instantly notices your frustrated demeanor. It reads slight levels of disgust and anger on your face, and decides to compensate for your possibly distracted mind in order to safely and quickly bring you to your destination: it will increase the frequency of its directional alerts and immediately alter its route prior to negative outbursts. The affect-mediated GPS would enable you to get to your destination despite those affective states which would impair your ability to plan and react soundly. Our research study shows that not only is this type of system possible but also many other affect-mediated creations that may help users complete their tasks.

Acknowledgements

The author would like to thank the director, staff, and children of the Carnegie Mellon University Children's School for their assistance with this work.

References

Chang, C.-C., and 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/%7Ecjlin/libsvm.

Csikszentmihalyi, M. 1990. *Flow: The Psychology of Optimal Experience*. Harper & Row, New York.

Gross, R.; Matthews, I.; Cohn, J.; Kanade, T.; and Baker, S. 2008. Multi-pie. In *Proceedings of the Eighth IEEE International Conference on Automatic Face and Gesture Recognition*.

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

Lucey, P.; Cohn, J.; Kanade, T.; Saragih, J.; Ambadar, Z.; and Matthews, I. 2010. The extended cohn-kanade dataset (ck+): A complete expression dataset for action unit and emotion-specified expression. In *Proceedings of the Third International Workshop on CVPR for Human Communicative Behavior Analysis (CVPR4HB 2010)*, 94–101.

Pai, V. 2012. Improving task performance in an affect-mediated computing system. Undergraduate honors thesis, Carnegie Mellon University. Paper available at http://search.library.cmu.edu/vufind/Record/1471396.

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

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