Automated Posture Analysis For Detecting Learner's Affective State

by

Selene Atenea Mota Toledo

B.S., Electronic Systems Engineering
Instituto Tecnológico y de Estudios Superiores de Monterrey
Cuernavaca, México
Mayo 1999

Submitted to the Program in Media Arts & Sciences, School of Architecture and Planning, in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE IN MEDIA ARTS AND SCIENCES

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2002

Author . . . .
Program in Media Arts and Sciences
August 9, 2002

Certified by . . . .
Rosalind W. Picard
Associate Professor of Media Arts and Sciences
Thesis Supervisor

Accepted by . . . .
Andrew B. Lippman
Chairperson
Department Committee on Graduate Students

Automated Posture Analysis For Detecting Learner's Affective State

by

Selene Atenea Mota Toledo

Submitted to the Program in Media Arts & Sciences, School of Architecture and Planning, in partial fulfillment of the requirements for the degree of MASTER OF SCIENCE IN MEDIA ARTS AND SCIENCES

Abstract

As means of improving the ability of the computer to respond in a way that facilitates a productive and enjoyable learning experience, this thesis proposes a system for the automated recognition and dynamical analysis of natural occurring postures when a child is working in a learning-computer situation.

Specifically, an experiment was conducted with 10 children between 8 and 11 years old to elicit natural occurring behaviors during a learning-computer task. Two studies were carried out; the first study reveals that 9 natural occurring postures are frequently repeated during the children's experiment; the second one shows that three teachers could reliably recognize 5 affective states (high interest, interest, low interest, taking a break and boredom).

Hence, a static posture recognition system that distinguishes the set of 9 postures was built. This system senses the postures using two matrices of pressure sensors mounted on the seat and back of a chair. The matrices capture the pressure body distribution of a person sitting on the chair. Using Gaussian Mixtures and feed-forward Neural Network algorithms, the system classifies the postures in real time. It achieves an overall accuracy of 87.6% when it is tested with children's postures that were not included in the training set.

Also, the children's posture sequences were dynamically analyzed using a Hidden Markov Model for representing each of the 5 affective states found by the teachers. As a result, only the affective states of high interest, low interest, and taking a break were recognized with an overall accuracy of 87% when tested with new postures sequences coming from children included in the training set. In contrast, when the system was tested with posture sequences coming from two subjects that were not included in the training set, it had an overall accuracy of 76%.

Thesis Supervisor: Rosalind W. Picard,
Associate Professor of Media Arts and Sciences

This research was supported by NSF ROLE grant 0087768 and also is an output from a research project funded by MIT Media Lab.

Automated Posture Analysis For Detecting Learner's Affective State

by

Selene Atenea Mota Toledo

The following people served as readers for this thesis:

Thesis Reader . . . .

Justine Cassell
Associate Professor of Media Arts and Sciences
MIT Media Laboratory

Thesis Reader . . . .

Deb Kumar Roy
Assistant Professor of Media Arts and Sciences
MIT Media Laboratory

Acknowledgments

First and foremost, I would like to thank to professor Rosalind Picard for being an incredible advisor and giving me the opportunity to joint her team at MIT. Thank you Roz for motivating me every day with your interesting and inspirational ideas, and being so charming and supportive; I felt very fortunate having you as an advisor personally and professionally because you have given me more than I could ever ask for.

Thanks to my readers Deb Roy and Justine Cassell for their advice and excellent reviews that made possible to shape my ideas.

I would like to thank to all the kids and their parents because without them this experiment could not be possible, and because to be bored is not pleasant. Also, for providing me invaluable feedback from teacher's perspectives, I thank to Kate, Dina, and Patricia who shared their time and ideas to help me understand a little bit more the art of teaching.

Special thanks to Stefan Agamanolis for his help with ISIS, Joel Stanfield from Steelcase for providing the Tekscan hardware system used in this thesis, and Hong Tan and Lynne Slivovsky for helping me to set up the chair hardware system and sharing with me their previous findings.

I am thankful to Linda Peterson for all her help and guidance during my stay at the Media Lab. In special at the end of this process in which I had to many things to do and short time for completing them, thanks Linda.

Also, I would like to thank to the LC group: Rob Really, Barry Kort, Cati Vauvacelle, and Tim Bickmore for sharing many interesting thoughts and always answer my innumerable questions.

Thanks to the affective computing group: Carson Reynolds, Raúl Fernández, Yuan Qi, Karen Liu, Stoffel Kunen, and Ashish Kapoor for being a group of great and talent people that stimulate my imagination with wonderful ideas and so interesting and fascinating conversations. My special thanks to Ashish Kapoor for being a wonderful friend and officemate for the last two years and have the patience to listen and answer my endless questions.

Thanks to Telmex, Carlos Slim, and Javier Elguea for their initiative to invest in research and for giving me the opportunity to come to MIT and demonstrate the high quality of Mexican Students. Also, I would like to thank to my ITESM professors: Dr. Jose Torres and Dr. Carlos Gomes-Mont for trusting on me and motivating me when I needed most.

Thanks "Pepetl" for your friendship and for making MIT welcome and friendly place. In special, I thank to my incredible, dearest, and unconditional MIT friends Raul Blasquez-Fernandez, Jose Maria Gonzalez, Ricardo Garcia (Ragito) and Alexander Stouffs. Thanks guys for sharing with me wonderful moments, helping me when I needed most and giving me the opportunity to have your wonderful friendship.

Finally, thank you to my family, especially to my dad, my dearest sister Ari, and Robert. Thank you for giving me so much love, for giving me the wings that let me fly and explore the world. Thank you for always believing in me, even when I did not. Thank you for always being close to me, despite the physical distance. Thank you for caring for me, everywhere, anywhere, always.

Papá, Ari, y Robert:

No tengo palabras para expresar todo lo que ustedes han hecho por mí. Gracias por todo su amor y apoyo cada día. Quiero expresarles que todo lo que hago es por ustedes y para ustedes, les dedico esta tesis con todo mi cariño.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

In human-human communication, body language is an excellent source of information. Its relevance roots from the assumption that it expresses implicit but true feelings. Consequently it is commonly assumed that the information from the nonverbal sources is more valid, more truthful and more revealing [1]. In the context of learning, non-verbal behavior can give us valuable information about students' affective states ([2],[53]).

Psychological literature and empirical experiments made by experts in the field of non-verbal language [10,11,16] present evidence that some postural behaviors are correlated with the level of interest and the degree of agreement or disagreement to the topic.

Based on these evidences, in a learning scenario where the child is interacting with the computer there are lots of interesting questions that we can ask. For example: Is the postural behavior that occurs in a face-to-face conversation similar to that occurring in a human computer interaction? Do children behave in the same manner when they are learning while interacting with humans as compared to when they are learning while interacting with a computer? How well can the theories that are based on studies made with adults be applied to the children?

Bull [3] presents results about the correlation between the affective states of interest and boredom with static postural behavior. Smith in her master thesis [4] correlates static body position (forward vs. backward) with the level of attention or engagement during human and human-like agent interaction. The study of Cassell, Nakano et al. ([18], [54]) did look at dynamic postural behavior, specifically they looked *when* people shift from one posture to another, and *how long* posture shifts lasted, and correlated it to the introduction of a new topic. Hence, the work developed in this thesis looks at the correlations over time of the static body positions focusing at a learning situation. Also, this work is looking at human-computer engagement and how postural behavior correlates with affect.

Specifically, one of the applications for which this research has been motivated is the Learning Companion Project (LC). This project aims to create an affective peer/tutor system that assists young students during a learning task. A critical part of the Learning Companion is, that it has, as one of its components, the interpretation and recognition of postural behaviors, the component on which this thesis is focused. As a result, the scenario of the recognition system is constrained to that of one child solving a mathematical puzzle in front a computer and the range of affective states are restricted to those that can be present in learning-computer situations [5]. It is important to highlight this thesis is focusing on the same context that the LC, but it doesn't mean the Learning Companion needs to use it, in special, because the ability to infer information about the learning situation from postures is something that along this thesis will be investigated. A detailed explanation of this project is showed in the applications section.

1.1 Thesis Objectives

The research questions in this thesis are as follows:

1. Do patterns of postural behaviors correlate with *some* of the affective states that occur naturally when a child is working in front of the computer trying to solve a puzzle ?

Further,

2. Can the computer -using two matrices of pressure sensors mounted on a chair- recognize those patterns?
3. Can a Hidden Markov Model be used to detect those posture patterns?
4. Can the posture patterns be differentiated across different children?

As I mention before there are several studies that suggest that there are some correlation between postures and states of the mind of speakers [3,4,18,54,15,16]. Hence, for the reason that there are not a clear articulation between postures and learning, this thesis tries to find experimental support about the relationship between postures and some of the affective states occurring in learning. It explores whether a set of static postures exists that can -trough their causal dependencies- detect in an automatic way dynamic postural behaviors correlated with affective states associated to interest and boredom.

Another objective is the implementation of a system that can be able to classify in real time the static postures made by a child during the learning-computer interaction. In particular, two matrices of 42 x 48 pressure cells made by Tekscan [6] were chosen as the posture-sensing device. These matrices are mounted on the seat and the back of a chair, and once an individual is sitting, the pressure sensors capture the body pressure distribution. Hence, the implementation of the static posture recognition system involves developing a technique for extracting and classifying reliably the body pressure features.

Finally, the last objective is to analyze the static posture sequences using Hidden Markov Models [7] and to examine if there are some patterns correlated with the investigated affective states.

I should mention that in this thesis I am not focusing on the distinction between *affective states* and *cognitive states* because even though, in the learning research community there are an extensive discussion about the difference between these two concepts, the aim in this thesis is to recognize behaviors that are highly correlated with the states themselves (called cognitive or affective) and not the study of their conceptual interpretation.

1.2 Applications in Computer Human Interaction

1.2.1 Learning Companion

The learning companion aims to be a computerized system sensitive to the affective aspects of learning which facilitates the child's own efforts at learning [8]. Learning the complex ideas involved in science, math, engineering and technology and developing the cognitive reasoning skills that these areas demand often involve failure and a host of associated affective responses. These affective responses can range from feelings of interest and excitement to feelings of frustration and boredom. The student might quit if he is not able to recover from the 'feeling of getting stuck'.
Skilled humans can assess emotional signals with varying degrees of accuracy, and researchers are beginning to make progress giving computers similar abilities at recognizing affective expressions. Computer assessments of a learner's emotional/cognitive state can be used to influence how and when an automated computational agent chooses to intervene.

The Learning Companion aims to sense surface level behaviors that are highly correlated with emotional and cognitive aspects during the learning-computer interaction in an unobtrusive manner. Hence, a critical part of the system performance is to develop mechanisms to sense the surface level behaviors without interfering with the natural learning process.

	On Task	*Off Task*
Posture	Leaning Forward, Sitting Upright	Slumping on the Chair, fidgeting
Eye-Gaze	Looking towards the problem	Looking everywhere else
Facial Expressions	Eyes Tightening, Widening, Raising Eyebrows, Smile	Lowering Eyebrow, Nose Wrinkling, Depressing lower lip corner
Head Nod/ Head Shake	Up-Down Head Nod	Sideways Head Shake
Hand Movement	Typing, clicking mouse	Hands not on mouse/keyboard

Table 1-1 A "common sense" list of Surface Level Behaviors

In general, teachers have told that cues like facial expression, eye gaze, hand gesture and posture help expert them to recognize whether the learner is on-task or off-task. These surface level behaviors and their mappings are loosely summarized in table 1. Whether all of these are empirical important, and are the right ones remains to be evaluated, and it will no doubt take many investigations. Such a set of behaviors may be culturally different and will likely vary with developmental age as well. The point is that exist a variety of surface level behaviors related to inferring the affective state of the user, while he or she is engaged in natural learning situations. This work is only one part of the learning companion system and focuses on analyzing the postural behaviors of a single-child solving a mathematical puzzle in front a computer.

1.2.2 Other Applications

Automatic analysis and understanding of body postures associated with affective states may potentially be used in designing virtual classrooms, in which machines can be aware of the user's affective state and, therefore, try to respond to them in an appropriate manner similar to human responses. Another possibility is one in which the machine can inform a human in a remote location, perhaps in a distance learning situation, about the affective state of the student or students. Software agents may also potentially use the output of the system to decide on effective communication strategies, or even to synthesize their own postural changes consistent with interest or boredom, since the models we use for posture analysis can also be used for postural synthesis. Automatic posture analysis will also have widespread application in psychological studies of non-verbal communication. In automotive applications, the system can be used for inferring information about the driver's behavior. It is also possible that the tools developed here may be of use in analyzing behavior related to seating comfort.

1.3 Outline of the thesis

The system of body posture understanding is divided into two problems: human analysis and machine posture analysis. The human analysis handles the problem of finding which are some of the affective states that are correlated with the children's postural behavior based on teachers' assessments. The machine posture analysis addresses the problem of developing an automatic system that can analyze the children's posture patterns based on the labels provided by the teachers in the human analysis part.

In summary, the organization of this thesis is as follows:

- Chapter 2 reviews previous research in human postural behavior, emotions, and learning. Also, this chapter gives an overview of related research in automatic posture recognition and interpretation.

- Chapter 3 describes the experiment that was conducted with children for collecting data from naturally elicited behaviors associated with interest and boredom, as well as, the method for establishing the ground truth about the affective states of the children during the computer interaction. Furthermore, it explains the experiment carried out for establishing the appropriated set of static postures that best describes the naturally gathered children's posture data.

- Chapter 4 explains the overall system for the automated recognition of posture patters associated with the affective states found in the experiment.

- Chapter 5 summarizes the results and concludes the thesis with suggestions for future work.

Chapter 2
Background

This chapter describes relevant theories of human postural behavior. Given that this thesis work focuses on the dynamic modeling of postural behaviors that are highly correlated with some of the affective states that a child is having during a learning-computer interaction, this chapter emphasizes particularly those theories that deal with the interpretation of non-verbal behavior, emotions, and learning. First, studies that support that non-verbal cues can be used to infer a student's affective states during a learning task are presented. Second, specific studies about posture interpretation are addressed. Third, previous systems for automatic posture detection and/or interpretation are described.

Essentially, the issues raised by this chapter are the most relevant ideas that form the foundation of this thesis. However, more specific theories that are necessary for the comprehension of some design issues of this thesis are addressed in the beginning of each chapter.

2.1 Relevance of Non-Verbal Cues For Inferring Humans' Affective States

Ekman and Friesen [9] introduced the conceptualization of *non-verbal leakage* that is caused by differential controllability of the communication channels. In other words, it means when people try to conceal negative affect and transmit positive affect instead, their deceit might be more successful in controllable channels (speech content, face) and unsuccessful in less controllable channels (body, filtered speech). According to Ekman [1], this is based in the hypothesis that precisely because of the greater repertoire of facial movement, people may be more careful to control their facial movements when trying to deceive others and hence are more likely to give themselves away inadvertently through the body movements.

Specifically, a study carried out by Allen and Atkinson [11], Goldin-Meadow, S., D. Wein, et al. [62], and Goldin-Meadow, S., M. W. Alibali, et al. [63] show some empirical evidences that non-verbal cues can be used to indicate whether a student is understanding a lesson.

Ekman in his early work ([1], [12], [13]) argued that people make greater use of the face than the body in judgments of emotion, that their judgments are more accurate when made from the face and that they can reach greater agreement in judging the face. At one stage, Ekman [14] proposed that the face is perceived as carrying information primarily about what emotion is being experienced, whereas the body is perceived as conveying information about intensity of emotion. Subsequently Ekman and Friesen [13] proposed that stationary facial expressions and postures are more likely to convey gross affect (such as liking), whereas movements of the face and body are more likely to convey specific emotions. Nevertheless, Bull [15] based on several experiments, presented results showing that both movements and positions convey information about four distinctive emotions and attitudes (interest / boredom, agreement / disagreement), and hence, contrasting with Ekman, he proposed that posture does constitute a significant source of information about people emotions and attitudes.

2.2 Posture Behavior Interpretation

Mehrabian and Friar [16] conducted several experiments where American male and female students were asked to think they were conversing with someone, and to adopt the positions they would employ to convey different attitudes while seated. From these studies, a number of postures have been the particular subject of investigation, namely, trunk lean forward, backward and sideways, body orientation, arms akimbo and body openness. Their findings revels that people believe that leaning forward or a decrease of leaning backward indicates a positive attitude. While sideways lean was found to vary according with the sex of both the message sender and the receiver. In the case of male encoders, intense dislike of another male was indicated by lack of sideways lean, whereas intense dislike of female encoders was indicated through greater sideways lean. In relation to status, sideways lean was used more when addressing someone of lower status. Mehrabian and Friar's studies didn't show clear results about body orientation. Observations of the arms akimbo position suggest that it has a generally negative meaning and it was used meanly by standing encoders. Observation of body openness (absence of folded arms or crossed leg positions) suggested generally positive meaning.

More recently, Rich at al. [17] in Mitsubishi Labs have defined symbolic postures that convey a specific meaning about the actions of a user sitting in an office which are: interested, bored, thinking, seated, relaxed, defensive, or confident.

In the field of non-verbal cues for discourse structure, Cassell, Nakano, & Bickmore ([18], [54]) at the Media Laboratory have also been conducting a study in which they provide empirical support for the relationship between postures shifts and discourse structures. They have found that postural shifts may be signal *boundaries* of units of information.

2.3. Systems For Automatic Posture Detection and/or Interpretation

Smith in her master thesis [4] created an interactive story-eliciting system for grandparents called *GrandChair System.* This system is based on a model of face-to-face conversation; tellers sit in a rocking chair and tell stories with the assistance of a conversational agent on a screen, who takes the form of a child, to help them tailor stories to a child audience, and prompts them with stories, questions, and video clips from their previous interactions. In particular, the system uses the combination of an accelerometer and a cushion sensor -resistive based sensor that provides information about the overall amount of pressure applied on it- for detecting grandparents' two major postures (forward and backward) or between rocking and not rocking motion; they used these changes in grandparents' postures to determine when a story was about to end. This system analyzes static but not dynamic correlations between those postures and the state of mind of the user. However, some of the disadvantages of this system are: First, before each session, the system needs to have a cumbersome calibration process. Second, it doesn't detect very reliably the postures when the user is moving constantly. And third, the couch sensor signal becomes invalid after the user has been sitting for a while, due to the fact that the resistive foam compresses.

Tan at all [19] proposed a system called the *Sensing Chair.* This system uses matrices of pressure sensors (2 of 42 by 48) fabricated by Tekscan [6] placed on the seat and back of a chair for detecting a set of predefined postures made by an user in an office environment. In her first approach, Tan classified the set of static postures using *PCA* (Principal Component Analysis). Specifically, Tan used a data set composed of 5 samples of 10 different postures made under command by 30 adult subjects. Using training and testing sets of different posture samples coming from the same subjects, she reported results of around 96 percent of posture recognition accuracy.

Later, Slivovsky and Tan [20] extended the *Sensing Chair* classificatory system to subjects that the system had not seen before (Multi-User recognition). In this work, the training and testing data sets contained posture samples coming from different subjects. However, using only *PCA*, they reported that the recognition rate of static postures went down (around 79 per cent). As a consequence, in order to improve the posture recognition rate, the original system was modified into a two-stage classification system, using either a Bayesian Classifier or one that uses a pyramid representation. Hence, the overall recognition rate increased to approximately 84 percent correct. Also, the Sensing Chair system was extended to classify in real time between static and transitional postures.

The mean differences between the system developed by Slivovsky and Tan and the work of this thesis are: First, In this work the system has been testing and trained on continuous postures made by children in a natural situation, whereas, the Slivovsky at al.'s system was trained with specific postures made under command by adult subjects. Second, the algorithm for feature extraction proposed in this thesis (see section 4.2) is thought to exploit the geometrical properties of the posture pressure maps, and it is different from the PCA technique used by Slivovsky at al. Furthermore, testing with the children's data base, the algorithm used in this thesis showed the advantage of modeling the pressure posture data better than PCA, recognizing very well the new children's postures. Finally, in this thesis, dynamic posture classification is developed to recognize postural behavior that is highly correlated with *some* of the affective states presented in a child's learning-computer interaction. In contrast, Slivovsky at al. are focused on distinguishing only between static and transitional postures.

Chapter 3

Human Analysis: Data Collection and Human Coding

Data collection for affective studies is a challenging task. We need to elicit affective states, like interest and boredom, on demand, which is almost guaranteed not to genuinely bring out the required emotional state. The subject needs to be exposed to the conditions that can elicit the required emotional state in an authentic way.

There are several research methods that can be used for eliciting and studying emotions. Particularly, in the learning and education literature, the issue of which is the appropriate context that can be used as elicitors of natural responses has been long debated among educational researchers representing different scientific disciplines (see [21]).

The *educationalists* (e.g. [22]) prefer to apply research methods in classroom settings. To investigate the classroom phenomena, they advocate the use of various ecological approaches that consider all relevant characteristics of the classroom ([23], [24]).

The *psychologists* follow a more rigorous experimental approach, which requires isolation of variables and control of *external noise*. Thus, they would like to eliminate any contextual variation and conduct their investigation under laboratory-like conditions, so that empirical causality can confidently be attributed to particular variables.

The methodologies of evidence followed for each of these two approaches are sufficiently contrasting to make *evidence* obtained under one perspective unacceptable to followers of the other perspective.

In this study we conduct the experiment in an ecologically valid setting, but controlling the external variables. This choice increases the confidence in the likelihood of obtaining empirical causality between the variables.

Another challenge for this study is to get as much information as we can about the *true* affective states that the child is experiencing. There are several ways to try to infer the true emotional state. These can be by self-report, measurement of biological signals or by observing verbal or nonverbal cues.
Self-report of emotions has the drawback that humans are notoriously bad at assessing how they feel ([25]). As a consequence, self-report in the field of emotions has been long known to be inaccurate in social science research (see [26] for a broad discussion of this topic).

The measurement of biological signals has the advantage of accessing uncontrollable changes that the body undergoes while experiencing the emotions. Nevertheless, in this study we decided not to use biological sensors, since during the first few experiments, we observed that the sensors tend to be uncomfortable for the children. As a result the children are likely to modify their behaviors and experience distraction, and adding a considerable amount of noise to the experiment. Also, I believe that biological sensors are probably another way of correlating with true emotional state, but don't yet allow us to infer it directly.

Instead of the methods mentioned above, we decided to focus on observation of non-verbal behavior. Behavior, in particular, acquires its relevance from the assumption that it expresses implicit but true feelings, that it taps the underlying affective layer and exposes attitudes and emotions that are hidden or even intentionally cancelled ([2]). In particular, studies made by LeDoux [27] and Damasio [28], show that important elements of human emotion are non-cognitive and emotions can affect action in ways that the person often cannot explain. Hence in this study, we analyze the non-verbal behaviors made by children and their correlation with the judgment of teachers about the children's affective states.

In this chapter, the first section describes the experiment with the children for eliciting natural responses associated with interest and boredom. The second section explains why the study needed a Structured Observation methodology [31] for establishing the ground truth about the affective states of the children during the computer interaction. The third section presents the details about the pilot study for getting the appropriated affective states to be used. The fourth section presents the details about the coding study made with teachers for assessing the children's affective states. Finally, the fourth section presents the study for establishing the set of static postures to be used by the automated posture system.

3.1 Data Collection: Children's Experiment

3.1.1 Apparatus

In order to elicit natural behaviors, the space where the experiment took place was a naturalistic setting (a common area called "The Cube" located at the MIT Media Lab building) and was arranged with the special chair (with the pressure sensors), a computer with a 21" inch monitor, mouse and keyboard on a normal table. To this space we have added three cameras, one pointing below the monitor directly upwards to the eyes (Blue

Eyes Camera), the other on top of the monitor capturing the facial expressions, and the last one on a small tripod at the side recording the posture image. We made the cameras less visible to encourage more natural responses. In total, we gathered 3 channels of video, one through a Blue Eyes Camera [29], two through Sony EVI-D30 Pan/Tilt/Zoom cameras.

We recorded five sets of data:

- The sensor chair pressure patterns
- Sony video-camera capturing the posture
- IBM Blue Eyes video of face
- Sony video-camera capturing the frontal face
- Computer Screen Activity

The data from the chair, the cameras and the computer program were synchronized.

As mentioned previously, a naturalistic setting was chosen to achieve ecological validity, since the study intention was to examine complete natural behaviors as close as possible to the way a child behaves in a computer-learning situation. Furthermore, this validity plays an important role on the effectiveness of the Structured Observation (see section 3.3.3), given it will influence the human coders' affective states interpretation.

Figure 3-1 Experiment Space.

The chosen method of recording employed in the experiment was aimed to be both unobtrusive and to preserve as much as possible the original behavior. An unobtrusive measure will minimize the effect of the observer on the subjects; there is little point in having a perfect technical record of behavior that lacks ecological validity because the participants have altered their behavior as a consequence of being observed. The type of video cameras employed was small in size and the appearance is different from the traditional video cameras. Also, the sensor chair is a non-obtrusive sensor that can provide the best of both worlds; it helps to preserve the original behavior without modifying the ecology of the environment of the participants.

3.1.2 Game Description

Fripple Place [30] is a constraint satisfaction game in which children try to match resident creatures to their assigned rooms. It explores three activity areas: (1) deductive and inductive reasoning, (2) synthesis and analysis of information to select options and form conclusions, (3) interpretation of evidence and predicting outcomes.

Figure 3-2 Fripple Place interface

This game was chosen because it is easy to learn - it can take from two to three games- and at the lowest level each game takes about 5 minutes. In particular, with this game, there is a high probability than it will soon elicit negative affective states like boredom or fatigue. In other words, this game could elicit both interest and boredom during 20 minutes of interaction.

3.1.3 Subjects

Contacting the subjects

Subjects were recruited via fliers posted on 10 public Cambridge and a private Newton (Montessori) Elementary School official boards with permission of each school principal. The flier suggested "*Win a fun tour to Media-Lab.*"; followed by, "*It's that simple. Come to the MIT Media Lab and test a new educational computer game.*"

Subject's parents voluntarily responded to the flier by contacting the experimenters, via contact information (phone and email) printed on rip-off stubs at the bottom of each flier.

Parents responding by email were then sent an email back, giving broad details of the educational game and their child's participation in it (according to the flier), including the information about the MIT Media Lab tour, along with a request for scheduling, and suggested times for coming in. Parents responding by phone were given a verbal version of the same material.

Parents were briefed about the game and the nature of the study. The children were told that we wanted to know how fun, friendly and interesting the game is.

Subjects

In total, 25 subjects came to participate in the study. In addition, two were not considered due to problems with the sensors, two more were not taken into consideration, as they needed to go to the restroom in the middle of the interaction. Another subject was not considered as she accidentally closed the game screen. One subject was already familiar with the game and had been playing the game in the past. Also, there were three subjects who played the *Incredible Machine* [61], which was very interesting for the children, and during the time of the interaction, never could bore them. As a result, data gathered from 16 children ages 8 to 11 years old was considered to be in good condition and without any problem.

To synchronize and manipulate five channels of data is a very complicated and time-consuming task. Due to the time constraints and considering the enormous amount of work the teachers would be doing while coding the data, we consider a subset of 10 children (5 male and 5 female) ages 8 to 11 years. As previously mentioned, every subject came from either a Cambridge or a Newton school. In fact, they were probably from relatively affluent areas of the state, although both schools integrate students from a variety of cultural and economic backgrounds. The subjects took part in the experiment one by one and none of them had played the game before.

3.1.4 Procedure

Before the structured interaction began, the experimenter introduced herself. The experimenter asked, in an informal way, the subject's name, school grade and 3 general questions regarding the participant interest in games or computers. After this, the experimenter showed the *Fripples Place* game to the participant and gave to him general instructions about how to play it (for more information about the game see appendix A).

The participant was asked to play the game once, ask any questions he had about the game, and subsequently, he was instructed to play alone. The interactions with the computer were videotaped and at the end of each session, subjects and their parents were informed that the subject had been videotaped and permission was requested to use the tapes for research. In no case this permission refused.

Each channel of information was synchronized. Every video frame from the face and the posture, each pressure distribution matrix from the chair and the game status were labeled with a time stamp. The computers and camera clocks were synchronized in order to assign the same time stamp to different device data. The time window for the time stamps was in the order of milliseconds.

3.2 Finding the Ground Truth: Assigning the children's Affective States

In behavior interpretation models there is always the issue of defining the ground truth - the *true* affective state of the learner. In the case of this experiment, to decide each child's affective state, 3 expert teachers were asked to label the children's video sequences. The three teachers were required to provide at least one label per minute of video and to indicate whenever they noticed a behavior that made them think a specific affective state was occurring (see section 3.4.2). This Structured Observation methodology [31] for human labeling was applied because the experiment was conducted with children between 8 and 11 years of age: for children this age, it is not reliable to give them questionnaires or interviews after 10 minutes of interaction. Normally they are not able to explain what happened regarding affective state changes during the session, particularly because children at this age have little understanding of the emotional language. This assumption was made in consultation with Dr. Jerome Kagan, an expert in such experiments ([32]).

3.2.1 Method Specifications

Given that we cannot directly observe the student's internal thoughts and feelings, nor can children of this age range reliably articulate their feelings, we choose instead to focus on labeling behaviors that communicate affect outwardly to an adult observer.

The coders were asked to perform a systematic observation looking for the following affective states: *Interest, Neutral, Taking a Break, Other and I do not know.* They were asked to provide comments for every *other* label they found, in which they must specify the name of the affective state, a brief description about the behavior that made them think about that other label and why (for details see section 3.3.4).

Note also that these five labels were arrived at after several iterations with pilot coders, as we tried to hone in on a set of relevant states that had reliably observable behaviors for the data we collected.

3.2.2 Software for coding

In order to aid the process of coding the data, I implemented a coding system based on the ISIS language (see [33]). The system is characterized by its ability to: (i) reproduce and annotate video streams frame-by-frame or at 3 different speeds (slow, normal and fast motion); (ii) go to a specific point in the video; (iii) save all the annotated data to a text file; (iv) generate or read a file that contains annotated data; (v) visualize the annotations on the screen, allowing an easy location of the areas of interest associated with a specific label.

The program allows coders to view and browse the video containing a segment of behavior several times. This way, the same behavior can be observed at a number of different levels and it permits coders to concentrate, on different occasions, on different aspects of the behavior. It also allows coding time stamps, as well as, durations of certain behaviors. Another advantage is that the video can be played at different speeds so that a behavior, which occurs for a very short duration, can be detected. The program also allows the reliability of measures to be checked more easily.

Figure 3-3 Isis program screen with tags indicators.

Command	Action
>	Frame forward
<	Frame Backwards
Space Bar	Change between Slow (10 fps) and normal speed (30 fps)
Drag mouse on "Movie Bar"	Goes forward over the whole movie locating it where the vertical yellow bar indicates.
Click with the mouse on "Movie Bar"	Goes to specific frame in the movie
Click on Label	Mark the start point of the label in the part of the video is being played.

Table 3-1. Screen Video Controls.

3.3 Pilot Coding Study

This study was an empirical first approach for exploring if *some* affective states could be reliably detected based only on a side angle video of a learner's posture, omitting game status and direct facial views. Three coders (MIT graduate students, 2 women and 1 man), were asked to label the children's posture videos with one of the followings categories: *interested, thinking, taking a break, confused, neutral, other, bored, distracted, tired, frustrated.*

Figure 3-4 Coding screen presented to no-teacher coders in the pilot study

As previously mentioned, one of the main differences between this coding and the coding performed by the teachers, is that the graduate students were watching the postural behavior without the face and the game status information. In addition, these coders were looking at a bigger set of affective states, 10 categories in total. Figure 3.4 shows the coding screen presented to this set of observers.

The pilot study was done using 10 children videos; each video was split in three segments of approximately 7 minutes long. As result, we had 30 different video segments that were chosen randomly for forming three sets of samples that were used for coding different rounds.

Figure 3.5 presents the total probability distribution that each coder assigned to each affective state for the entire video: *interested, thinking, taking a break, confused, neutral, other, bored, tired, distracted, and frustrated.* Table 3.2 shows the coders' Kappa results.

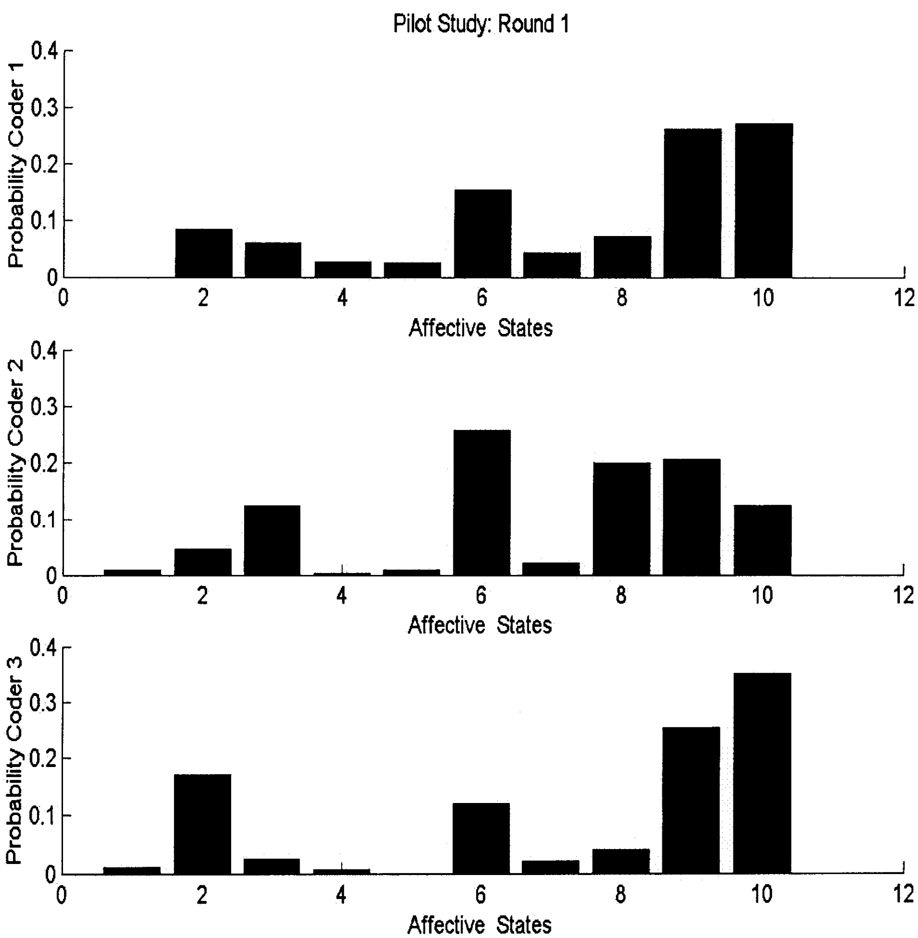


Figure 3-5 Probabilities obtained for each coder after the first round of the pilot study. The numbers represents the Affective States as follow: 1. Frustrated; 2. Bored; 3. Tired; 4. Distracted; 5. Other; 6. Neutral; 7. Confused; 8. Taking A Break; 9. Thinking; 10. Interested.

From the results, it is evident that the level of agreement between coders was very low. Therefore, in order to refine the coding for the main experiment, it was necessary that the coders get together and discuss their differences. Appendix C gives an explanation of the Cohen's Kappa formula and section 3.3.4 gives details about which is an acceptable level of agreement.

Kappa Round 1	Value
Coder1 and Coder 2	0.3926
Coder 1 and Coder 3	0.2995
Coder 2 and Coder 3	0.3115

Table 3-2 Cohen's Kappa calculation for measuring the level of agreement between the non-expert coders in the first round.

In this first round of the pilot study, the coders pointed out they could not distinguish between the *interest* and *thinking* classes. For example, if the child was interested, then probably she was also thinking about the problem. Alternatively, if he was thinking about something else besides the problem, then probably he was not only distracted, but also daydreaming. As a consequence, the *interest* and *thinking* classes were combined into one class. This new class was still called *interest*, but emphasizing the fact that the student is interested only when the student is thinking about the problem.

Similarly the classes of *bored* and *tired* were confused during the first round. The coders thought that when the child is getting tired, she tends to get bored as well. Similarly, when the child gets bored he starts to get tired. We defined a new class called *bored*, which occurred when the child was tired and stopped working on the task altogether. The class did not include the case when the child was tired, but still putting a lot of effort in trying to solve the puzzle.

After practicing with some examples, the coders coded for a second time. Figure 3.6 shows the final probability distribution for the different affective states for the second round. The tables 3.3 to 3.6 show the Cohen's Kappa and confusion matrices for this second round.

Kappa Round 2	Value
Coder1 and Coder 2	0.7136
Coder 1 and Coder 3	0.7715
Coder 2 and Coder 3	0.7283

Table 3-3. Cohen's Kappa calculation for measuring the level of agreement between the coders (no teachers) Round 2

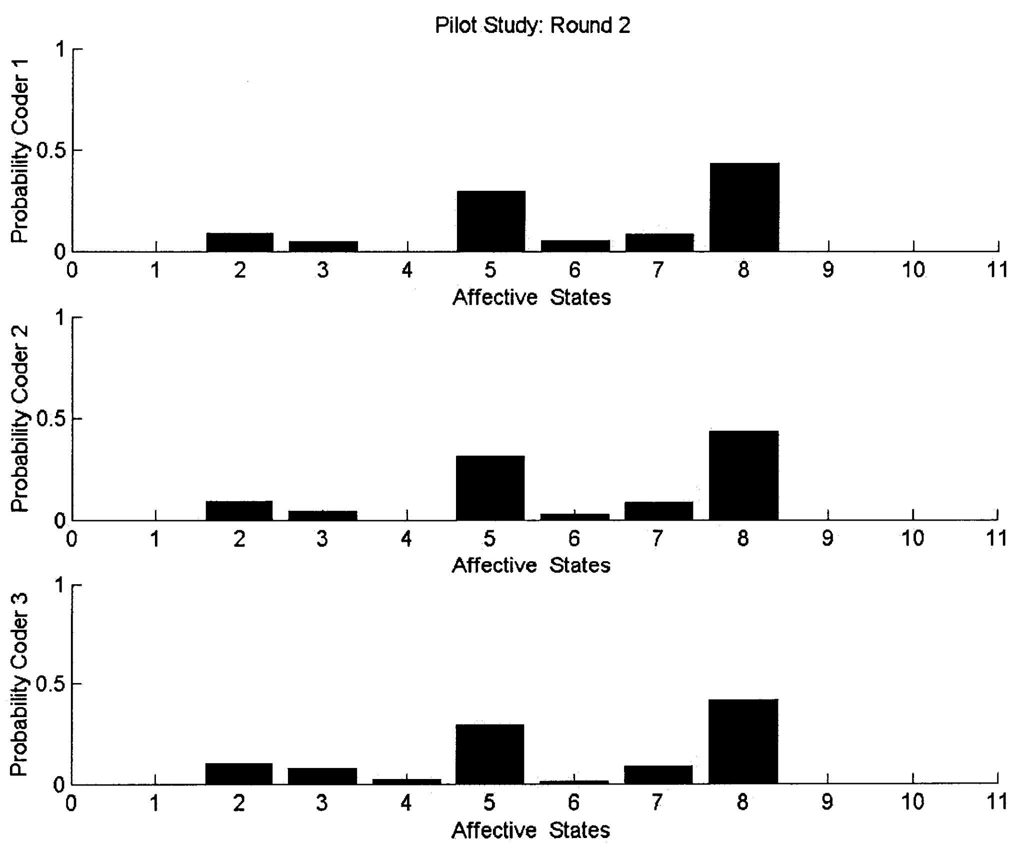


Figure 3-6 Probabilities obtained for each coder after the second round of the pilot study. The numbers represents the Affective States as follow: 1.Frustrated; 2.Bored; 3.Distracted; 4.Other; 5.Neutral; 6.Confused; 7.Taking A Break; 8. Interested.

	Frustrated	Bored	Distracted	Other	Neutral	Confused	Break	Interested
Frustrated	0	0	0	0	0	0	0	0
Bored	0	186	0	0	12	0	4	11
Distracted	0	46	2	0	22	0	9	37
Other	0	0	0	0	0	0	0	0
Neutral	0	1	53	0	643	15	0	28
Confused	0	0	20	0	81	0	3	32
Break	0	0	0	0	3	0	200	8
Interested	0	1	26	0	28	59	2	968
Class Agreement	0.0	0.7949	0.0198	0.0	0.8150	0.0	0.9174	0.8930

Table 3-4. Confusion matrix between coder 1 and coder 2, round 2.

	Frustrated	Bored	Distracted	Other	Neutral	Confused	Break	Interested
Frustrated	0	0	0	0	0	0	0	0
Bored	0	155	1	35	15	0	7	0
Distracted	0	1	88	4	0	0	0	23
Other	0	0	0	0	0	0	0	0
Neutral	0	2	58	15	654	0	6	5
Confused	0	76	25	0	3	12	0	20
Break	0	4	0	0	0	5	198	4
Interested	0	12	15	0	54	12	6	985
Class Agreement	0.0	0.6200	0.4706	0.0	0.9008	0.4138	0.9124	0.9499

Table 3-5. Confusion matrix between coder 1 and coder 3, round 2.

	Frustrated	Bored	Distracted	Other	Neutral	Confused	Break	Interested
Frustrated	0	0	0	0	0	0	0	0
Bored	0	177	49	0	0	0	0	8
Distracted	0	70	22	9	0	0	0	0
Other	0	0	0	0	0	0	0	0
Neutral	0	1	76	1	633	27	10	41
Confused	0	1	22	0	51	0	0	0
Break	0	0	13	0	0	0	202	3
Interested	0	1	5	44	42	2	5	985
Class Agreement	0.0	0.7080	0.1176	0.0	0.8719	0.0	0.9309	0.9499

Table 3-6 Confusion matrix between coder 2 and coder 3, round 2.

The results show that *interested, taking a break, neutral and bored* happened most often and had the highest agreement.

The objective of this pilot study was to improve the internal validity of the experiment, and find which affective state labels are valid and reliable. For validity, I wanted to understand the relationship between a *specific* affective state and the postural behavior. For getting a better reliability, I carefully defined the concept of each affective state. I tried to consider affective states categories, which are mutually exclusive (see section 3.4.2 for details about the description of each category). Finally, because I am employing a very subjective measure, I considered it necessary to be confident that the affective state categories are both valid and reliable.

3.4 Teachers' Coding

3.4.1 Revising Teachers' agreement

Once I finished devising the affective states classificatory system, I prepared three video examples for each category: *interest*, *taking a break*, *neutral* and *bored*; 12 examples in total. Each instance was approximately 60 seconds long and included the video for the face, posture and game status. Subsequently, three female expert teachers were trained in its use, and using new episodes of behavior, they were prompted to code 80 new instances, where each of them was approximately 60 seconds long. The instances were extracted from each of the ten children videos with two segments of each of the four categories per child. Each teacher was asked to assign only one label to every example. The examples were presented to the teachers in a random order one by one.

Then, using the data gathered from the teachers' labels, the inter-rater-reliability –that is, the degree to which raters, working separately, agree over their classification of the affective states- was checked using Cohen's Kappa formula [34] (see Appendix C for details of this formula). The results for the agreement between coders were the followings:

Kappa Teachers	Value
Teacher 1 and Teacher 2	0.7800
Teacher 1 and Teacher 3	0.8432
Teacher 2 and Teacher 3	0.7339

Table 3-7 Cohen's Kappa coefficients for the teachers' study

	High Interest	Interested	Neutral	Taking A Break	Bored	Other	I do not know
High Interest	13	2	0	0	0	0	0
Interested	0	14	1	6	1	0	0
Neutral	0	0	16	2	0	0	0
Taking Break	0	2	0	12	0	0	0
Bored	0	0	0	0	11	0	0
Other	0	0	0	0	0	0	0
I do not know	0	0	0	0	0	0	0

Table 3-8 Confusion matrix between teacher 1 and teacher 2

	High Interest	Interested	Neutral	Taking A Break	Bored	Other	I do not know
High Interest	14	1	0	0	0	0	0
Interested	0	17	2	2	0	1	0
Neutral	3	0	14	1	0	0	0
Taking Break	0	0	0	14	0	0	0
Bored	0	0	0	0	11	0	0
Other	0	0	0	0	0	0	0
I do not know	0	0	0	0	0	0	0

Table 3-9. Confusion matrix between teacher 1 and teacher 3

	High Interest	Interested	Neutral	Taking A Break	Bored	Other	I do not know
High Interest	12	1	0	0	0	0	0
Interested	2	13	0	3	0	0	0
Neutral	2	0	14	1	0	0	0
Taking Break	1	3	2	13	0	1	0
Bored	0	1	0	0	11	0	0
Other	0	0	0	0	0	0	0
I do not know	0	0	0	0	0	0	0

Table 3-10. Confusion matrix between teacher 2 and teacher 3

From the results presented above and according to Robson C. [48], who reports that Kappa in the range 0.4 to 0.6 is fair, between 0.6 and 0.75 is good and above 0.75 is excellent, I evaluated the level of agreement between teachers good enough for continuing to code the complete set of data.

3.4.2 Coding the complete set of data

After the level of reliability was assessed and before teachers started coding, the experimenter provided them with a written description of every affective state category: *interest, taking a break, neutral, bored, other and I don't know*, (see table 3-11 for each category description). As I mentioned before, the selection of each label was based on the validity and reliability analysis presented in section 3.3. In the case that the teachers assigned the category o*ther*, they were instructed to specify the affective state name, as well as a description based on the behavior observed.

It is important to highlight that the teachers were not aware of the final purpose of the experiment, which was to find the correlation between postural behavior and affective states. However, they knew the purpose of the experiment was to correlate affective states with behavior. That means they were also observing behaviors that do not necessarily correspond to posture, for example face and arm gestures.

In total, 200 minutes of video were scored, around 20 minutes per child. Teachers scored two children per session; each session was, on average, 2.5 hours long and was realized over different days. The sequence of every child's video was chosen in a random way.

Class	Definition
Interested	1. When the student is attending to or performing the task. 2. When the student is thinking about the problem. 3. Does not include the case when the student is thinking about something else besides the problem. Then, probably, the student is distracted or daydreaming.
Taking A Break	4. When the child has been attending but after some time he or she starts to move around just refreshing her or his body, but quickly the student comes back to the task.
Neutral	5. When the student doesn't show any affective state in specific, but is still involved in the learning task.
Bored	6. When the kid is not interested in the task. 7. When the child was tired and stopped working on the task altogether. 8. Does not include the case when the student is tired, but she is still putting a lot of effort in the learning task.
Other	9. When the affective state observed doesn't involve any of the categories mentioned above and the teacher can specifically identify the affective state.
I Do Not Know	10. When the affective state observed doesn't involve any of the categories mentioned above, but the teacher cannot identify the affective state.

Table 3-11. Affective States Descriptions

3.4.3 Teachers 's Coding Results

The original categories chosen for the affective states were: *Interest, neutral, taking a break, bored, other, and I don't know.* However, during the study with the teachers the following two observations were made:

Observation 1: During the process of checking the coder's reliability of the affective states, the class *neutral* was successfully differentiated from *interest* (see tables 3.8 to 3.10). However, when the teachers started coding the complete interaction data, they consistently marked under the *other* category the distinction of different levels of interest. Specifically, *high interest, interest,* and *neutral* were classified consistently as three different levels of interest. Hence, these three classes are interpreted in the rest of this thesis as *high, medium, and low interest.*

Observation 2: There were some affective states like *distracted, confused, puzzled,* and *satisfied,* which were annotated consistently under the *other* category. These states were not considered, as the notes made by the teachers suggested that the children's facial expressions made them interpret the affective state they coded.

3.5 Establishing the Basic Set of Postures

3.5.1 Method

There are several criteria for classifying postures ([35], [36], [37]), but the main difficulty in all these approaches is that they divide the body movements in terms of several units (head, neck, legs upper, legs lower, shoulders, trunk upper, waist, etc.). In these systems, it is not possible to describe *leaning forward* as a single behavioral unit; instead, the basic postures are describe in terms of several positions (trunk, head, neck, shoulders upright, legs upper and waist straight, legs down touching floor, and after, trunk forward, head upper 20 degrees more than the previous one, upper legs straight, forward waist with trunk with frontal view, down legs slightly behind the chair, etc.) and hence the basic unit of the movement is lost.

Peter E. Bull ([3], [38]), proposed a scoring system called *Body Movement Scoring System*, for movements maintained for at least one second. In this system postures are classified into four main types: head, trunk, arms/hands, and legs/feet, which occur in a face-to-face conversation between two persons. It might seem to be a good system for describing gestures, but it has not been widely used, as it is difficult to automate the classification of the movements. Usually there are human coders that transcribe the movement descriptions.

This thesis follows the philosophy proposed by Bull in his *Body Scoring System*, as it uses movements rather than positions of the body parts as the basic unit of analysis; hence it is possible to describe postures as a series of movements rather than as a series of positions, capturing the natural structure of body movement.

3.5.2 Posture Coding Study

Two human coders (MIT graduate students) were trained to recognize the target postures based on both the posture and frontal face children videos (we use both videos in order to provide more information).

The coders' level of agreement was evaluating using two data sets. Each data set had 100 different video segments randomly extracted from the 10 children's posture videos. Each segment was 10 seconds long.

Both coders labeled the first data set with the following categories: *leaning forward, leaning backward, sitting right, sitting left, sitting upright, slumping, and other.* However, the level of agreement was low, 69 percent. The experiment was then discussed with the coders in order to determine how to improve reliability and accuracy.

Subsequently, the second data set was labeled as well, but adding the categories of sitting forward right, or sitting forward left, sitting backward right, sitting backward left, and sitting on the edge on the chair. And also, coders were asked to give a confidence level - low, high, medium- for each label. With these new definitions, the coder's agreement increased to 83 percent. Table 3.12 shows the second round posture categories.
Hence, only the children's posture samples with high and medium level of confidence were used for training the algorithm for the static posture recognition.

Sitting on the Edge
Leaning forward
Leaning forward right
Leaning forward left
Sitting Upright
Leaning back
Leaning back right
Leaning back left
Slumping Back

Table 3-12 Set of static posture categories

3.6 Chapter Summary

1. Posture video, frontal video, game screen video and the observations from the sensor chair were recorded for 10 children, while they interacted with the computer.

2. Three expert teachers coded the children's videos; for the affective state they thought each child was having during the computer learning interaction.

3. From the teachers' coding, we found teachers distinguished reliably the following affective states: *high*, *medium*, and *low interest*, *taking a break*, *bored*, and *other*. We did not consider the category *other* as the teachers mentioned in their observations that the other category correlated with the facial movements rather than the postures.

4. Two coders assessed the children's videos according with the postures they were observing. Hence, it was found that nine postures were frequently repeated during the experiment.

As a result, we have the following set of data: 200 minutes of video that have been labeled with 5 affective categories, which are synchronized with their corresponding body pressure distribution map, captured at a rate of 8 frames per second.

Chapter 4

Machine Analysis: System

In order to analyze postural behaviors, first it is necessary to specify the set of static postures to be recognized (see section 3.5). Hence, once the set of postures has been defined, the next step is to extract relevant features that best represent the posture data. The posture data is obtained using two matrices of pressure sensors made by Tekscan [6], which are mounted on the seat and back of a chair, and once a person is sitting, the matrices of sensors capture the body pressure distribution. Consecutively, the next task is to classify the extracted features for recognizing the static set of postures. Finally, the sequence of static postures over time is used for estimating the child's affective states during the computer-learning interaction. The performance of the affective recognition task depends not only how well the static postures are recognized, but also on how well the temporal patterns of these postures represent the affective states, as labeled by the human coders (see chapter 3 for details).

Most researchers have focused on the recognition of gestures made under command [39]. One of the algorithms used for recognizing postures using pressure sensors mounted on a chair is Principal Component Analysis [19]. Some other different approaches [4,57,38] have used accelerometers, or magnetic or light sensors for detecting some static postures during a face-to-face conversation. All of them have oriented their systems to adults.

This chapter explains a system for children that can combine both the automated posture recognition in real-time and the analysis of the postural behavior over time for estimating the children's affective states. The system is divided in three parts. Figure 4-1 gives you an overview of the system.

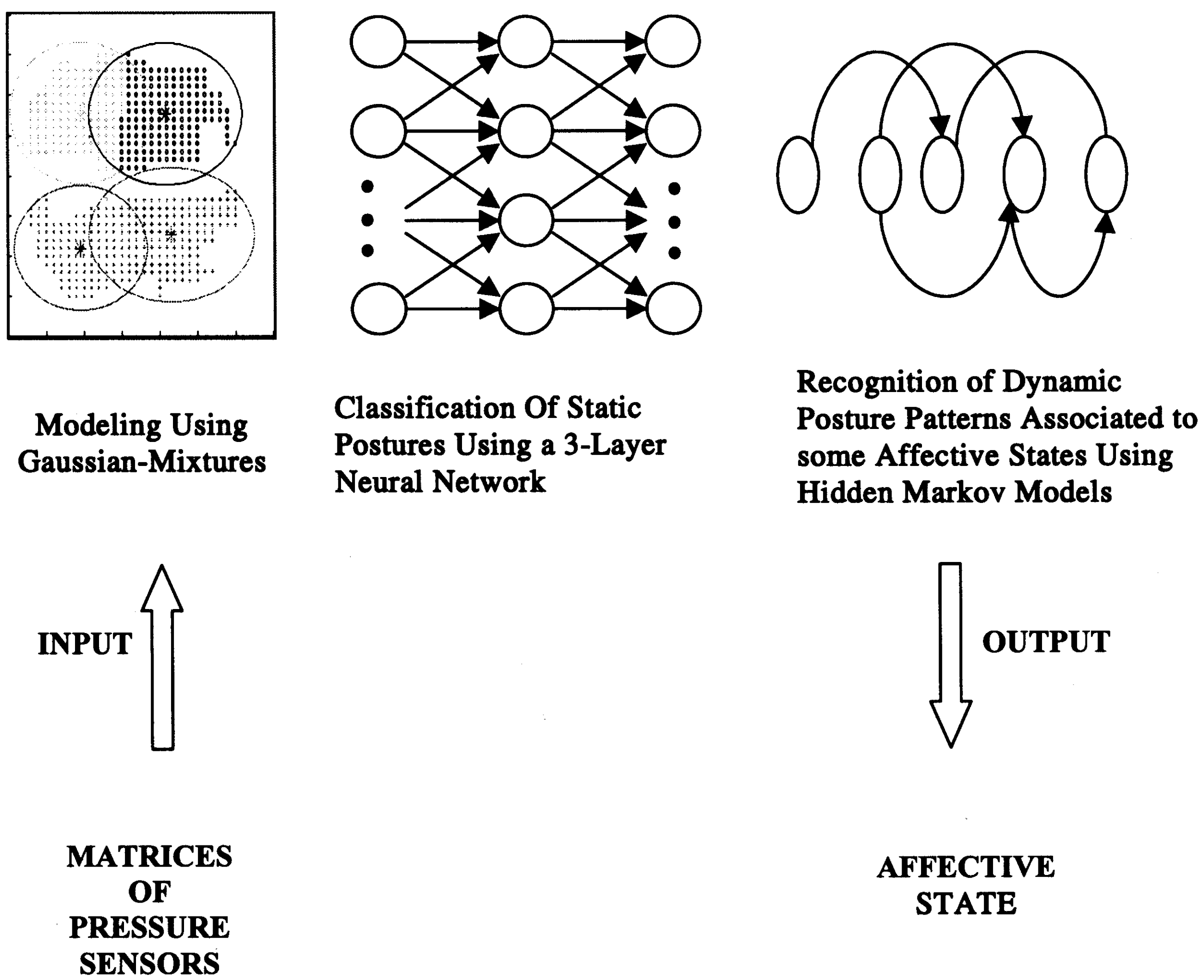


Figure 4-1. System Overview

The first part is concerned with the extraction of features that best represent the postures. The second explains, after the features are extracted, how these features are used for classifying the postures using a feed-forward neural network. Finally, the third part explains how the postural behaviors (over time) associated with some affective states are recognized using Hidden Markov Models.

It is relevant to keep in mind that the system has, as a basis, empirical results that were found through the experiments explained in chapter 3, where the basic set of static postures and the children's affective states to be recognized were established. This chapter explains the architecture of the overall machine analysis system in detail.

4.1 Sensor Selection

One of the challenges is to select the appropriate sensor for recognizing the postures. The sensor should be able to recognize postures across a wide variety of users and environments. It is very difficult to use a camera with computer vision techniques for the task as the variations in lighting, change in camera positions etc. can disrupt the posture recognition processes. We use *the sensing chair* to sense the posture. The sensing chair has pressure sensors mounted in a chair and was previously implemented by Hong Tan (see [19], [20]). This sensor is able to identify the posture as a single movement, and at the same time, is not obtrusive, which preserves the children's natural behaviors during the interaction.

4.2 Static Posture Recognition

4.2.1 Algorithm highlights

In this thesis, from the results presented in section 3.5 of this thesis, children's postures were classified in the following categories:

1. Sitting on the edge,
2. Leaning forward,
3. Leaning forward right,
4. Leaning forward left,
5. Sitting upright,
6. Leaning backwards,
7. Leaning back right,
8. Leaning back left,
9. Slumping back.

This set of postures doesn't include, for example, the category of crossing legs, as in our set of approximately 200 minutes of data; the children never crossed the legs. In contrast, the set includes sitting on the edge of the chair and leaning forward right and left, and leaning backwards right and left, that describe when the children make fast movements with legs, or move side to side on the chair.

For the reasons mentioned above, the algorithm has to have the following features:

1. It needs to be sensitive to translation, as it needs to distinguish between a leaning forward posture and sitting on the edge of the chair.
2. It should be very robust to the subject size: The chair will be used by children between 8 and 11 years old, and at this age the corporal size tends to have a big variation.

3. It should be robust to low-resolution data. This is important because the long-term goal is to develop a low cost posture sensor, and in order to do that, the resolution of each pressure sensor and the number of them are variables that add to the cost.

4.2.2 Hardware

The postures are recognized using a pressure sensor made by Tekscan [6] mounted on a chair. This sensor uses an array of force sensitive resistors and is similar to those used by Tan [19] and later by Slivovsky & Tan [20]. It consists of two 0.10 mm- thick sensor sheets, with an array of 42-by-48 sensing units with an overall area of 41 x 47 cm. Each unit is a variable resistor and the normal force being applied to its superficial area determines its resistance. This resistance is transformed to an 8-bit pressure reading. Hence, the level of pressure can be interpreted as an 8-bit grayscale value. As result, the pressure distribution maps can be visualized as a grayscale image. One of the sheets is placed on the backrest and one on the seat. The pressure distribution maps (2 of 42x48 points), according to the specifications, could *sense* at a sampling frequency of 127Hz. Figure 4-2 shows an example of the body pressure distribution matrices.

Another important point is to choose the type of chair appropriate for the task. In prior research, the sensors were tested with adults using a *Herman Miller Aeron Chair* [55]. We put special attention into choosing the chair, as it could modify children's behaviors, and as a result, we could have behaviors produced by the chair itself rather the underlying affective state.

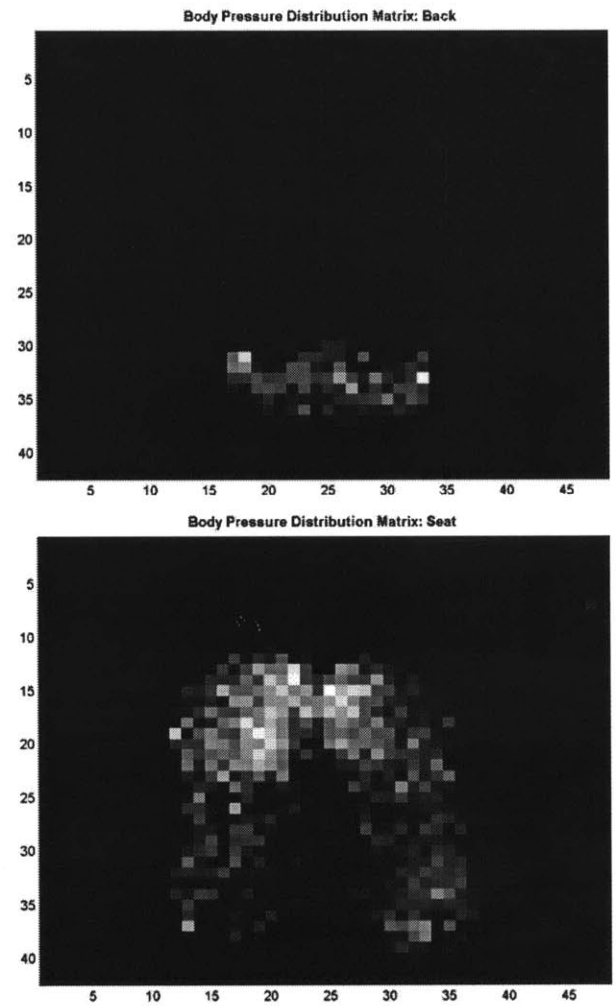


Figure 4-2 Example of the body pressure distribution matrices while a child is sitting upright on the chair; the figure below is the seat and the figure above is the back.

We found that studies made by Helander [41] present evidence that the posture that people express are a reflection of their feelings rather than the ergonomics of the chair for itself. And the constraint for considering these results is that the chair must not break the basic general requirements (relative size to the user and a seat-pan with comfortable curvature). As a result, we chose a Leap SteelCase chair [56] because it could be fixed to a wide range of sizes (seat pan and back rest altitude & openness) and it has a comfortable and firm curve seat. Then, in each experiment, we fixed the chair according to the subject size, by raising or lowering it. Figure 4-3 shows the chair with the two matrices of pressure sensors.

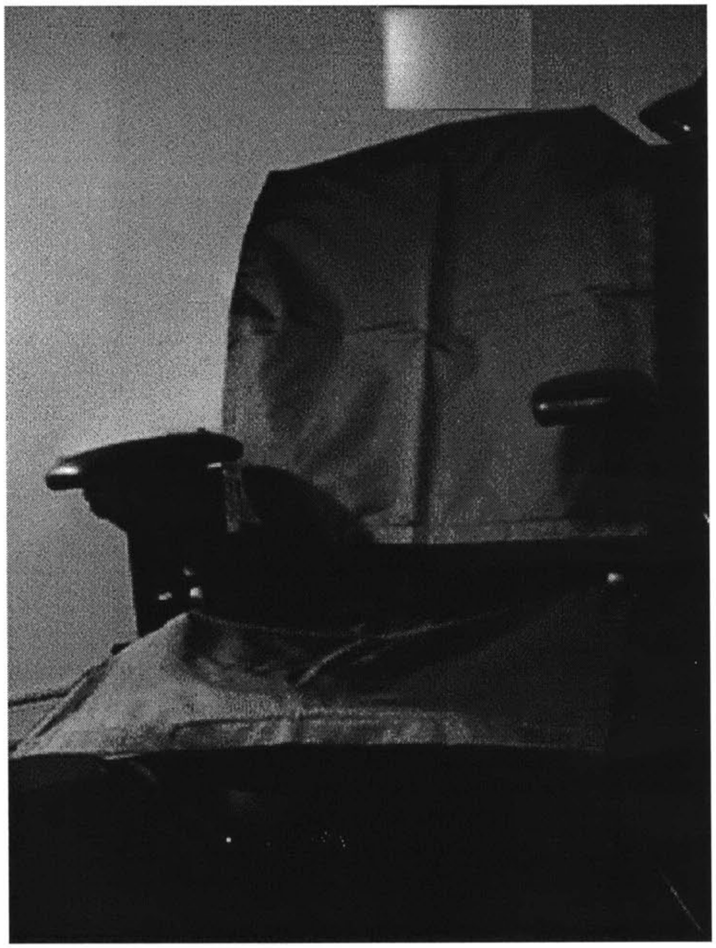


Figure 4-3 Chair with sensors

4.2.3 Software For Static Posture Acquisition

The pressure maps for the posture database were obtained from the experiment presented in chapter 3. The recording program was written in Microsoft Visual C++ 6.0 and it runs under windows 98. It uses the API library supplied by Tekscan [6] that permits the direct control and accesses to the pressure sensors interface board.

The program used in the experiment had a very simple interface. I recorded both pressure distribution maps: seat and back, together with a time stamp obtained from the computer clock that has been synchronized with the other devices used in the experiment (game computer and cameras). The recording rate used on this program is significantly less than the specifications claim is possible: 8 posture frames per second.

4.2.4 Posture Data Modeling

In our original problem of detecting different postures from the pressure distribution maps, we can observe that the data have a geometrical structure, which changes when a different posture is made.

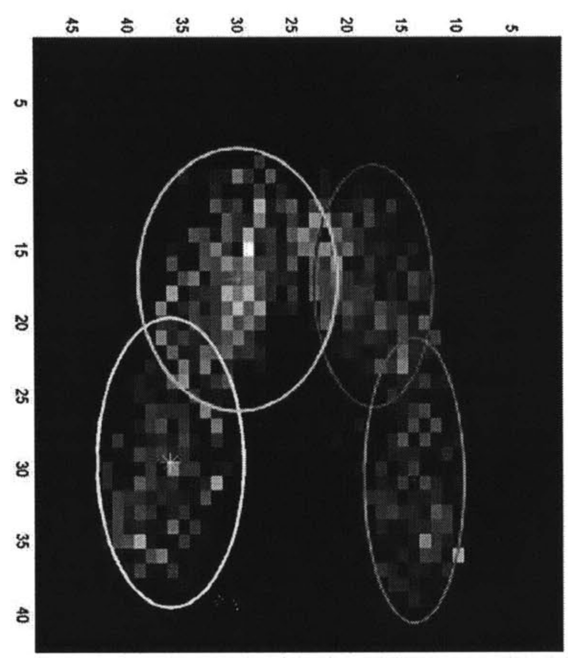


Figure 4-4 Seat pressure distribution matrix modeled with 4 gaussians. Each circle represents the parameters (mean and variance) of every gaussian.

This geometrical structure forms clouds of points in a 3-dimensional space. Suppose the points that form each cloud came from a single normal distribution. Then, its mean and covariance matrix gives the sufficient statistics of the data. In other words, these statistics constitute a compact description of the data. The mean locates the center in 3-dimensions of the cloud. It can be thought of as a single point that best represents all the data in the sense of minimizing the sum of squared distances from this point to each sample from each cloud. The covariance matrix gives a measurement of how well the mean describes the data in terms of the amount of spread that exists in various directions.

From above, assuming the sampled points come from a mixture of 4 normal distributions, (see figure 4-4) we can approximate the parameters of the mixture gaussians for describing the pressure data. In essence, the different sizes and orientations of the hyper ellipsoidal clouds can be used as features for classifying the different postures.

However, it is important to take into account that the problem of estimating the parameters of a mixture of gaussians is not trivial. Specially, it depends on the a priori knowledge of the data that determines which are the appropriate initialization points for the model. Erroneous initial parameters may lead to meaningless results and, instead of fitting the structure of the data; we would be imposing a structure on it.

4.4.5. Filtering the Raw Data: Noise Reduction

When the pressure sheets are located over the chair, they suffer small deformations on the edges of the pressure maps. In addition, the sensor itself adds some noise to the raw data; figure 4-5 presents a raw image of the pressure distribution map corresponding to a leaning forward posture. In order to eliminate the noise coming from the sheets deformation and the sensor, we applied two methods: a threshold function and a morphological operator.

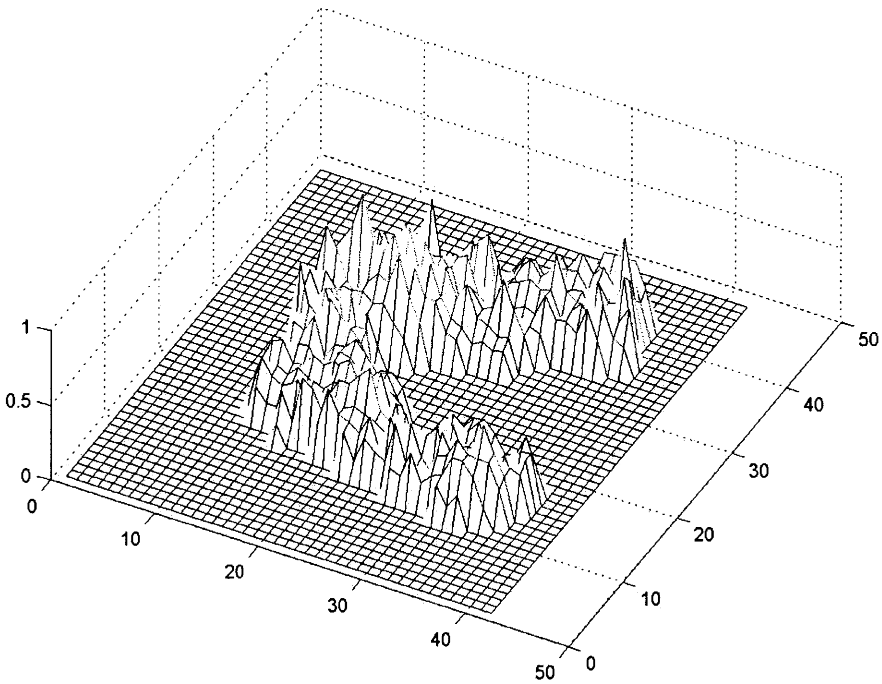


Figure 4-6 Seat Body Pressure Distribution Matrix after the noise was removed

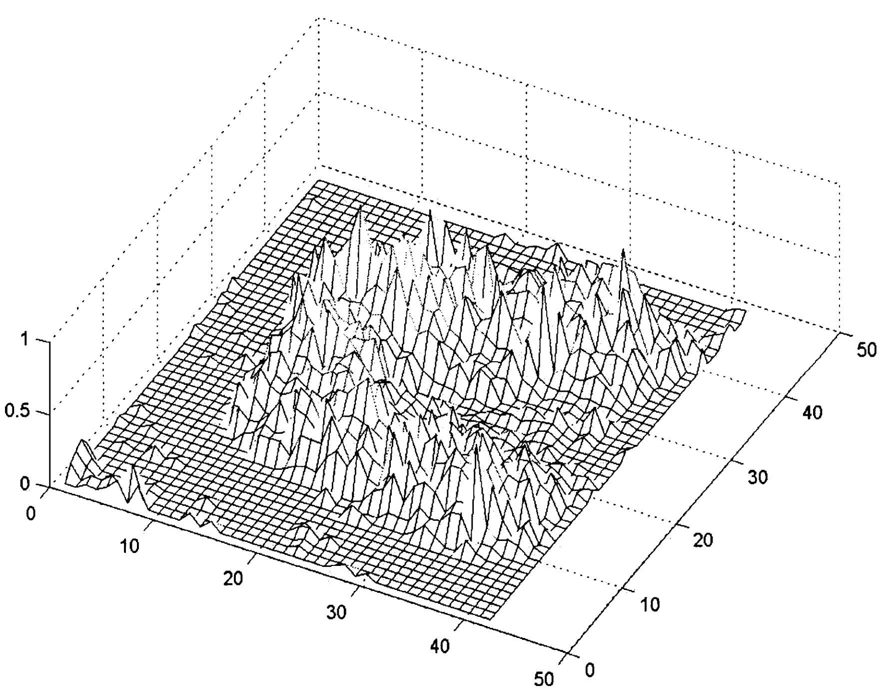


Figure 4-5 Raw Seat Body Pressure Distribution Matrix while a leaning forward posture is occurring

The threshold function was used mainly for masking out some pixels that belong to the pressure points caused by the sheets deformation. The function was applied to the raw data, taking as threshold value the ten percent of the maximum value of the pressure map.

Also, the morphological operation of erosion [42] was applied to the raw data. This operation was used to delimit the shape and boundaries of the body pressure distribution image and reduce the unwanted noise as well. Specifically, the erosion operation involves moving a kernel across the image. The kernel used is a simple square element with an anchor point of 3 (matrix 1-by-3). The operation is based in the condition in which a white pixel will remain white in the resulting binary image if all of its neighbors are white.

The mask resulting after applying the threshold and erosion functions cleans the raw data from two sources of noise. Figure 4-6 shows the seat pressure matrix presented in figure 4-5 after the noise was removed.

4.4.6. Modeling With The Expectation-Maximization algorithm

The expectation-maximization (EM) [43] algorithm was used to obtain the estimation of the four mixtures of gaussians that model the pressure sensor data. This algorithm is often used in estimation problems where some of the data are *missing*. In the posture pressure matrices, the missing data is knowledge of to which class each pressure point belongs.

In this application, the EM algorithm iteratively alternates between computing the lower bound (*E-Step*) and maximizing the bound (*M-Step*), until the point of zero gradient is reached. The *E-Step* and the *M-Step* appear in Equations 4-1 and 4-4 respectively.

Expectation Step:

$$\hat{P}(w_i \mid v_k, \hat{\theta}) = \frac{p(v_k \mid w_i, \hat{\theta}_i)}{\sum_{j=1}^{c} p(v_k \mid w_j, \hat{\theta}_j)\hat{P}(w_j)}$$ Equation 4-1

Maximization Step:

$$\hat{P}(w_i) = \frac{1}{N}\sum_{k=1}^{N} \hat{P}(w_i \mid v_k, \hat{\theta})$$ Equation 4-2

$$\hat{\mu}_i = \frac{\sum_{k=1}^{N} \hat{P}(w_i \mid v_k, \hat{\theta}) v_k}{\sum_{k=1}^{N} \hat{P}(w_i \mid v_k, \hat{\theta})}$$ Equation 4-3

$$\hat{\Sigma}_i = \frac{\sum_{k=1}^{N} \hat{P}(w_i \mid v_k, \hat{\theta})(v_k - \hat{\mu}_i)(v_k - \hat{\mu}_i)^t}{\sum_{k=1}^{N} \hat{P}(w_i \mid v_k, \hat{\theta}_i)}$$ Equation 4-4

$\hat{P}(w_i)$ is the a priori probability of gaussian i, $\hat{\theta}_i$ is the parameter vector estimated for gaussian i *where* $i = 1..4$, v_k is the 3-dimension pressure point *where* $k = 1..N$, and N is the number total of pressure points being classified.

Particularly in this application, the EM algorithm tries to find the parameters for the four gaussians that best represent the pressure data. It is trying to find the best natural grouping of data, finding to which cluster each pressure point has the highest probability to correspond. However, since EM only finds a local maximum, a good initialization is crucial.

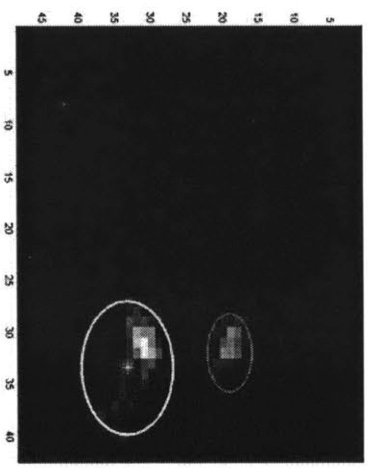
The EM algorithm was implemented with the following modifications:

1. The number of points for classifying is variable. After the data was cleaned, we took a reduced number of sub-sampled data points that have their original 8-bit pressure value.

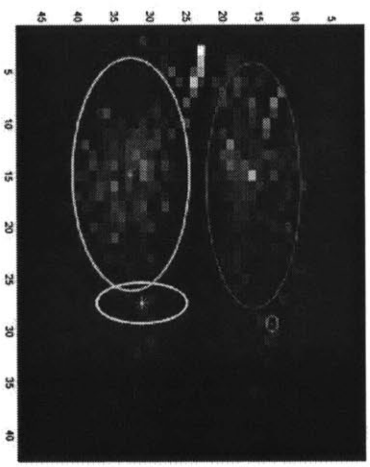
2. The maximum number of gaussians is fixed to four. This number was chosen, because after testing with several numbers of clusters and looking for some evidence of which is the best way to group the points in order to distinguish several postures, a distinctive pattern using four gaussians was observed; using four clusters in a geometrical representation, a posture is easily described and distinguished from others.

3. Another modification is that the algorithm is constrained to preserve the relative positions between the gaussians. For example, with the normal gaussian mixtures algorithm if you have just one leg on the chair, the four gaussians will be distributed on its area, and the algorithm might not be able to discern that is just one leg. In contrast, the modified algorithm can distinguish if just one leg is leaning on the chair. The side of the leg could be distinguished as well, but our approach is based on real data from learning experiments where children did not engage in a lot of unusual contact with the chair.

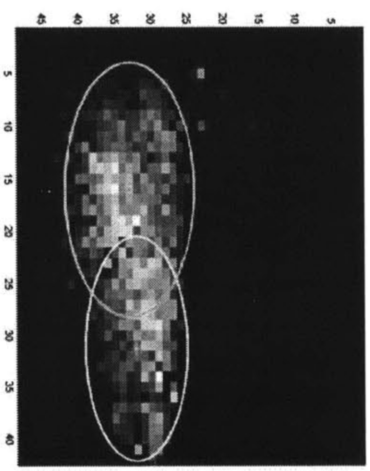
For the extraction of features we have been normalizing the data prior to applying the EM algorithm. This normalization just involves invariance of scale, rather than translation or rotation. Thus, all the features have unit variance, but not zero mean. This scaling has the advantage that the algorithm can distinguish between patterns of different postures (sitting transversally one side, leaning forward, and sitting on the edge) that for example a normal principal components algorithm cannot do. Figure 4-7 shows seven cases where the Gaussian Mixtures Algorithm has been applied.



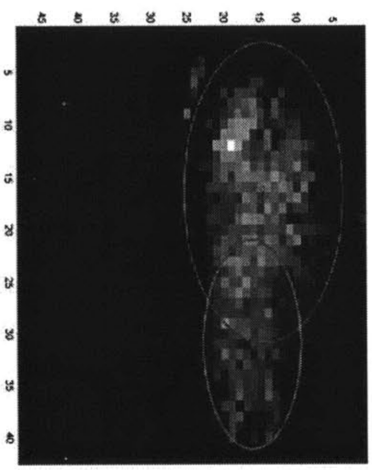
(a) Sitting On the Edge



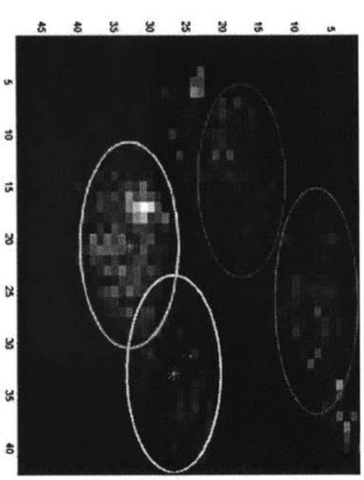
(b) Sitting Upright with knees raised



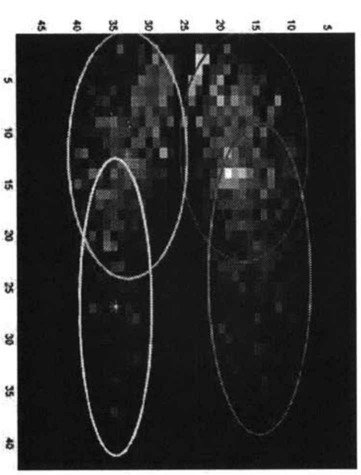
(c) Sitting on only right Leg



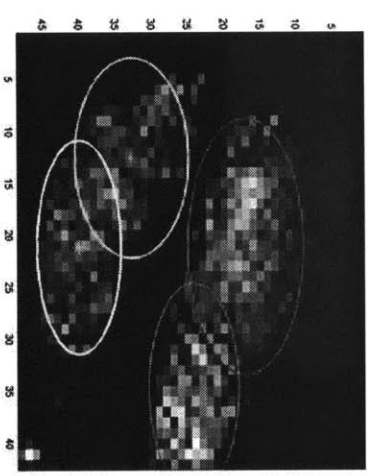
(d) Sitting on only left Leg



(a) Sitting transversely oriented to the right



(b) Sitting Upright



(a) Sitting transversely oriented to the left

Figure 4-7 Some of the features that the Gaussian Mixtures algorithm can identify. Each circle represents the parameters (mean and variance) of every gaussian.

4.3. Static Posture Classification

After the data features were extracted by modeling each pressure matrix with the Gaussian Mixture algorithm fixed with four gaussians, the parameters of each gaussian (mean and variance) are used to feed a 3-layer feed-forward neural network [44] that classifies the input data determining the static posture in real time.

In the next sections, I will give a short overview of neural networks, followed by the network architecture used for the posture classification based on the gaussian parameters estimated by the Expectation-Maximization algorithm (see previous section for more details). After that, I will explain how my data set was collected, as well as the specifications of the training parameters used. Finally, I will show the performance of the algorithm for recognizing the postures.

4.3.1 Short Overview Of a Neural Network

A neural network is composed of single units called *neuron* that are interconnected each to another. Figure 4-8 [44] shows the structure of a layer of neurons.

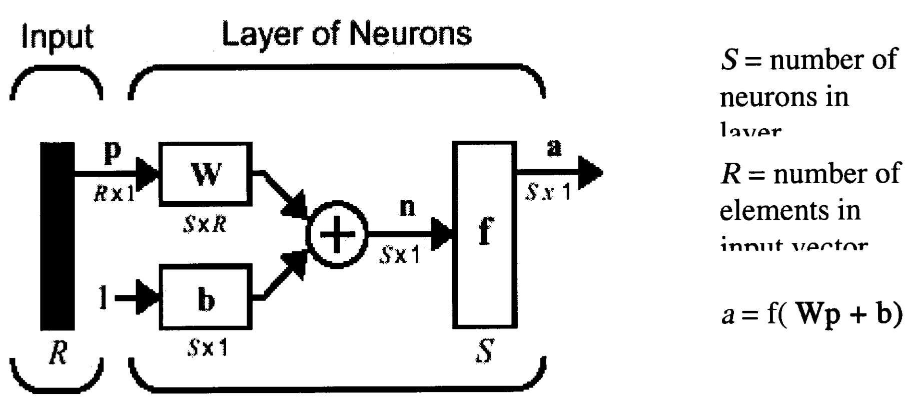


Figure 4-8 Structure of a Layer of Neurons

The scalar input p is transmitted through a connection that multiplies its strength by the scalar weight w, to form the product wp. Here the weighted input wp is the only argument of the transfer function f, which produces the output a. If the neuron has a bias, the net input is the sum of the weighted input wp and the bias b. This sum is the argument of the transfer function f. The transfer function f, typically is a step function or a sigmoid function, which takes the argument n and produces the output a. Note that the weight w and the bias b are both *adjustable* scalar parameters of the neuron [44].

Several neurons can be combined into multiple layers that have great power and flexibility, for example, feed-forward networks. The architecture of a multi-layer network is constrained, in part, by the problem to be solved. For example: the number of inputs to the network is constrained by the problem, and the number of neurons in the output layer is constrained by the number of outputs required by the problem. However, the number of layers between network inputs and the output layer and the sizes of the layers are up to the designer.

The central idea of neural networks is that their parameters can be adjusted so that the network exhibits some desired behavior. Thus, we can train the network to do a specific task, for example classification, by adjusting the weight or bias parameters to achieve some desired output.

4.3.2 Neural Network Architecture

Part of the architecture that was used in this application was partially imposed by the problem; we have the parameters of eight gaussians (four for each pressure matrix), each gaussian has in total 7 parameters formed by the x, y, and z values that locate the mean in the 3-dimensional space, the diagonal values corresponding to the variance, and one value corresponding to its prior probability. In total, for the eight gaussians we have 56 parameters that correspond to the neural network input vector. It should be noted that these gaussian parameters are mapped always in the same order to the neural network input vector.

As we want to classify a set of 9 postures based on the input vector, the size for the neural network output is fixed to 9 as well.

The rest of the neural network parameters were chosen trying to reach the highest performance in the recognition rate. Such parameters are presented as follow:

Type of Neural Network	Feed-forward back propagation with fully interconnected neurons in each layer
Size of Input Vector	56
Size of Output Vector	9
Training Function	Bayesian regularization algorithm
Performance Function	Mean Square Error

Table 4-1 Neural Network parameters

	Layer 1	Layer 2	Layer 3
Number of Neurons	56	12	9
Transfer Functions	Tan-Sigmoid	Log-Sigmoid	Linear

Table 4-2 Neural Network Layers

4.3.3. Data Set

Human coders were trained for coding the posture videos that were synchronized with the pressure distribution matrices recorded during the experiment described in chapter 3 (see section 3.5 for more details about the human posture classification). The classified pressure matrices were used as source data for posture recognition.

As I mention before, I am going to use data from body pressure distribution matrices (two matrices of 42 x 48 sensing points each) mounted on the chair. And I want to design a classifier that determines which of 9 postures a child sitting on the chair has.

For training the neural network, first, the data features were extracted using the Expectation Maximization algorithm, having as result a data set formed by vectors of 56x1 values, which are classified with one of the 9 postures that form the posture source data set.

Posture Class	Subject's Number									
	1	2	3	4	5	6	7	8	9	10
SE	50	66	0	0	257	24	0	52	97	0
LF	731	500	151	98	411	111	204	62	57	350
LF_L	264	663	18	21	415	399	471	142	18	231
LF_R	357	109	53	198	218	109	254	158	0	177
SU	267	400	417	932	384	441	12	689	656	485
LB	140	118	399	133	38	254	51	122	53	191
LB_L	120	143	150	54	14	47	318	15	39	162
LB_R	62	48	259	652	21	290	68	463	12	264
SB	82	115	32	74	0	56	26	29	0	51

Table 4-3 Details of the static posture data set. Leaning Forward (LF), Leaning Forward Left (LFL), Leaning Forward Left (LFR), Seating Upright (SU), Leaning Back (LB), Leaning Back Left (LBL), Leaning Back Right (LBR), Seating on the Edge (SE), Slumping Back (SB).

The posture data coming from 5 children (from child 1 to 5) were used as a training set, whereas the data coming from the other 5 (from child 6 to 10) were reserved for testing.

Particularly, I tried to balance between the training and testing sets cases in which children didn't have any example of some specific posture. Table 4-3 gives the details about the data set composition.

4.3.4. Neural Network Training Parameters

Postures coming from five different children in the database shown in table 4-3 contribute as the training set for the neural network. During training the weights and biases of the network were iteratively adjusted to minimize the network performance function. The performance function employed was the mean square error - the average squared error between the network outputs and the target outputs.

The training algorithm utilized was the Bayesian Regularization Algorithm [45]; its implementation was taken from the Matlab Neural Networks Toolbox [58]. This algorithm is a modification of the Levenberg-Marquardt [46] training algorithm designed to produce networks that generalize well and to reduce the difficulty of determining the optimum network architecture.

One problem that can occur when training neural networks is that the network can over fit on the training set and not generalize well to new data outside the training set. This problem was prevented by training with the Bayesian regularization algorithm and also testing several combinations of the number of neurons in the hidden layers according to the results of this algorithm.

Another problem is that caused by the error surface minima [47]. This causes nonlinear transfer functions in a multi-layer network to introduce many local minima in the error surface and, as gradient descent is performed on the error surface, it is possible for the network solution to become trapped in one of these local minima. As this may happen depending on the initial starting conditions, the weights of the network were initialized randomly several times to increase the chances that the neural network reached the best solution. In this thesis case, the network was trained until 2000 epochs or the squared error reach zero. Figure 4-9 summarizes the results of training the network using 2000 epochs.

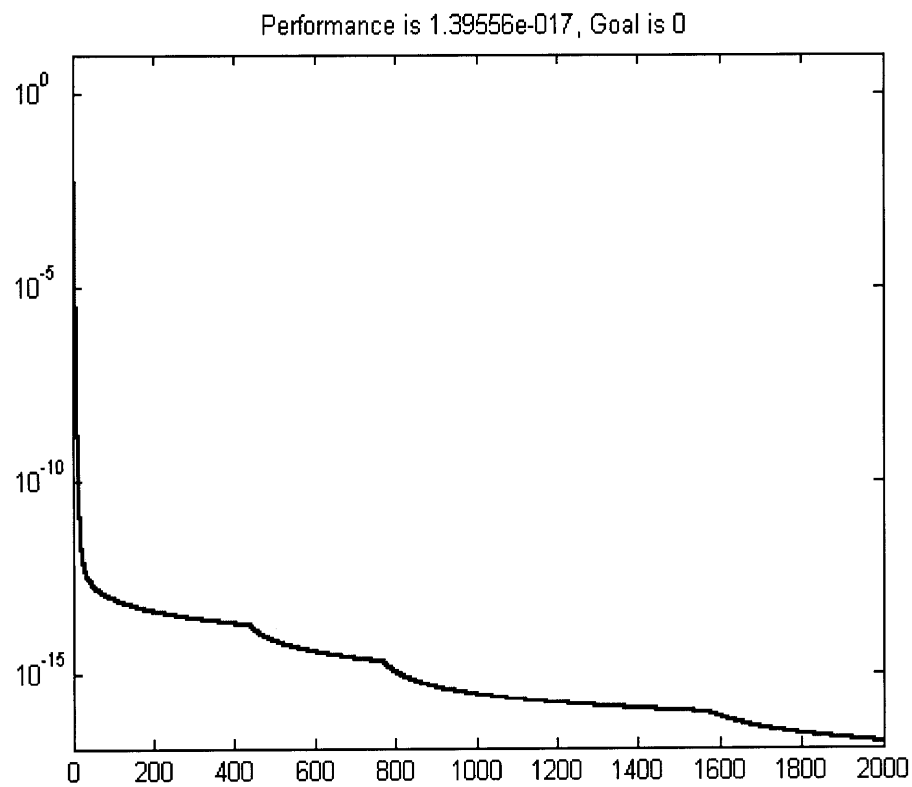


Figure 4-9 The convergence of the Neural Net during training using the parameters for which we got the best convergence

4.3.5 Classification Results

The Neural Network classification was tested using posture data of five new subjects from the database explained in section 4.3.3. Table 4-4 gives the confusion matrix obtained.

	Classification									
Data Set	**LF**	**LFL**	**LFR**	**SU**	**LB**	**LBL**	**LBR**	**SE**	**SB**	**% Recognition**
LF	758	26	0	0	0	0	0	0	0	96.68%
LFL	168	1009	44	0	0	0	0	40	0	80.02%
LFR	117	0	535	0	0	0	0	46	0	76.65%
SU	0	0	0	2128	84	0	0	0	71	93.21%
LB	0	0	0	61	610	0	0	0	0	90.91%
LBL	0	0	0	42	75	464	0	0	0	79.86%
LBR	0	0	0	22	58	0	981	0	36	89.43%
SE	13	1	0	0	0	0	0	159	0	91.91%
SB	0	0	0	11	0	5	0	0	146	90.12%
									Total	**87.64%**

Table 4-4 Confusion matrix and recognition rate of the Neural Network. Leaning Forward (LF), Leaning Forward Left (LFL), Leaning Forward Left (LFR), Seating Upright (SU), Leaning Back (LB), Leaning Back Left (LBL), Leaning Back Right (LBR), Seating on the Edge (SE), Slumping Back (SB).

The results show that the neural network can classify the postures with an overall accuracy of 87.64%. The classes of sitting on the edge of the chair, sitting upright, leaning forward, leaning back, and slumping back are classified very robustly. In contrast, the classification for the classes of leaning forward right and left and leaning back right and left is not as robust as obtained for the other postures; however, the recognition rates still range from 76.65% to 89.43% correct.

Notice that the data set employed has a considerable level of noise. We emphasize the samples in our data were obtained from natural made postures, which we think makes the problem harder than the detection from those postures made on command. Another remark is that the testing and training sets were built with different children; consequently, the neural network was classifying examples of postures that it had not seen before. Finally, it is relevant to highlight that the outcome of this static posture classification is used for feeding the inputs of different Hidden Markov Models (HMM) that estimates patterns of behaviors correlated with some affective states. The architecture of the Hidden Markov Models is explained in section 4.4.

Also, it should be notice that although the naturalness of the data, the algorithms, the age of the subjects, and the posture classes were completely different in our case, these results of an average 88% classification can be roughly compared to the results of Slivovsky and Tan [20], who obtained an average of 84% classification of ten static postures using the same matrices of pressure sensors.

Finally, it is relevant to emphasize that in this thesis a separate static classification stage was needed for two main reasons: First, when the HMMs were tested using the 56 gaussian parameters as emission probabilities, they were more susceptible to the variations of those parameters. Hence, even though it is not typically needed, a neural network between the gaussians parameters and the HMM was added. This neural network doesn't include any learning function, in other words, it is only used for giving to the HMM a simplified information about the user's posture. Second, in terms of behavior analysis, we also used the system as a tool to analyze how the dynamic patterns of behavior look like and compare these patterns with the results obtained by previous research in non-verbal behavior. Thus, the static classification stage allowed us to interpret and assess the dynamic posture information that was obtained by the experiment. It also permits synthesizing postural behaviors.

4.4 Dynamic Posture Classification

In this section I will present the third part of the system architecture and also, I will explain how the postural behaviors correlated with the set of affective states can be described as a dynamic model represented by *Markov Chains*. From this perspective I have assumed that affective states are considered to have characteristic posture sequences associated with them, each with its own interstate transition probabilities.

Specifically, this layer of the architecture uses a set of independent *Hidden Markov Models* (HMMs) [7] for recognizing the posture sequences. Each HMM input takes a sequence of classified postures; the classification of each posture is obtained from the feed-forward neural network output that constitutes the previous layer of the system architecture (See last section for details). The neural network output consists of an integer that identifies how the static posture was classified.

As an example of how this part of the system works, suppose that we have a sequence of classified postures made by a child sitting in front of the computer. We also have a *Hidden Markov Model* that represents every affective state. Next, each model computes the probability that the observed posture sequence was produced by it. Finally, the observed posture sequence is determined to belong to the model that has the highest probability.

Particularly, the set of affective states to be recognized are restricted to those that were classified by the teachers in the experiment stage of this research; these affective states are: *high, medium, and low interest, taking a break, and bored.* As I mention previously, the *other* category was not taken into account for two main reasons: First, the teachers indicated that most of the time they label with the *other* category the video data because of some facial expressions made by the child. Second, an increased number of affective states will require increased degrees of freedom –the number of models to be handled- by the overall system to adequately represent the complete set of affective states, which in turn may be too great to be meaningful and practical.

Finally, an overview of this section is as follows: Initially, I will present the notation and the training method that were chosen for modeling the affective states based on postural behavior. Afterwards, I will present the testing results followed by a discussion of them.

4.4.1. Notation

Let N represent the number of the hidden states in the model, $H = \{1 \ldots N\}$ denote the individual hidden states, and lower case h_t the hidden state at time t. Also, let $A = P_{ij}(h_t = i \mid h_{t-1} = j)$ be the hidden state transition probability matrix, which is the probability of transitioning from hidden state j to hidden state i.

In this thesis I am assuming the models are *Ergodic* and *Markovian*. *Ergodic* [7] means that every hidden state of the model could be reached from every other hidden state in the model in a finite number of steps. *Markovian* [49] means that given a number of states h_t in the past, where $1 \leq t \leq N-1$, only the most recent state h_{N-1} needs to be kept, as the earlier states provide no additional information useful in predicting the future state h_N.

K indicates the number of distinct posture symbols recognized by the model. These symbols correspond to the classified postures made by the child during the learning-computer interaction. Every posture symbol is represented by $V = \{v_i\}$, where $1 \leq i \leq K$. Whereas, O_t denotes an observation symbol from V in time t, where $1 \leq t \leq T$, and T denotes the number of observed postures in the whole posture sequence.

Additionally, let $B = P(O_t = v_i \mid h_t = j)$, where $1 \le i \le K$ and $1 \le j \le N$, denote the probability distribution of the observation symbol O_t, given the hidden state j ($h_t = j$) at time t. As well as, let $\pi = P[h_1 = j]$ be the initial hidden state distribution, where $1 \le j \le N$ and $t = 1$.

Summing up, the HMM for each affective state is fully determined through the set of observed posture symbols V, that were determined based on the experiment explained in section 3.5, along with the specification of the N and K parameters and the three probability matrices A, B, and π, which are unknown.

4.4.2. Model Selection

In this thesis, we don't initially know, which is the best model to use for recognizing the posture sequences associated with each affective state. For this reason, we will use the observed data –postures gathered in the experiment- not only for parameter estimation but also for model selection. This parameter estimation is related to the *learning problem* [7], and there are several methods that could be used to solve it. Those methods range from maximum-likelihood (ML), maximum-a-posteriori (MAP), or Bayesian methods, to more conventional techniques such as gradient descent, expectation-maximization (EM) [43], or the latest techniques such as maximum entropy discrimination [50]. In particular, I focus on estimating the ML parameters for each affective state model using the Baum-Welch algorithm [51] or E-step in the EM algorithm [43].

The training method of *k*-fold cross-validation ([59], [52]) was used for determining the model parameters N and T by choosing one of the several models that has the smallest generalization error. This method was implemented using Kevin Murphy's Matlab Hidden Markov Model Toolbox [60]. Specifically, having k equal 10; all the children's posture sequences were randomly divided in 10 sub-groups of approximately equal size. The model parameters were estimated 10 times, each time leaving out one of the subgroups from training, but using only the omitted subgroup to compute the chosen error criterion. In this thesis, I use the log likelihood [7] as the evaluation function; its equation is shown below.

$$\log P(O_t, H_t) = \log P(H_1) + \sum_{t=1}^{T} \log P(O_t \mid H_t) + \sum_{t=1}^{T} \log P(H_t \mid H_{t-1}) \qquad \text{Equation-5}$$

The figure 4-10 shows each model's generalization error for different values of hidden states (N). Hence, N's with the smallest generalization error and low variance were considered (Appendix B shows the model variances graphs). This initial computation of each model's generalization error used the sequences that were directly cut from the fragments where the three teachers agreed in their assessment (see section 3.4 for details). As a consequence, the sequences do not have a uniform length; they have T variable.

In a scenario were the computer needs to estimate the correlation between postural behaviors and affective states in real time, it is necessary to figure out which is the adequate sequence length the models will use to evaluate the observed postures.

Therefore, in this case, the appropriated length was determined computing the generalization error for combinations of different values of T, $T=\{8,16,24,32,40,64,88\}$, and, for each of these cases, the model with the three best values of N. Similarly to the computation of the generalization error for N, this calculation used the k-fold cross-validation method, with k = 10. Table 4-5 summarizes these generalization error results.

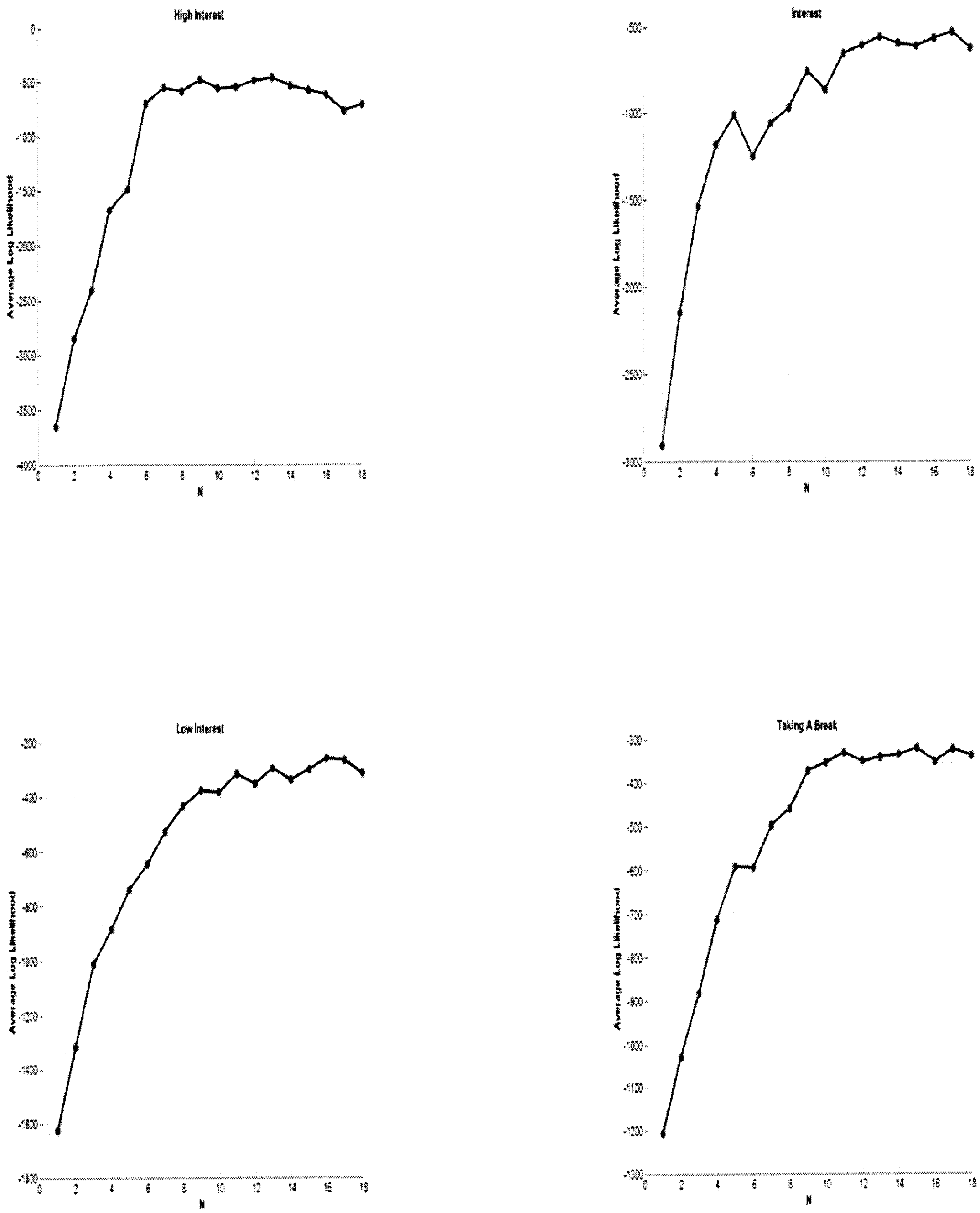


Figure 4-10 Model Generalization Error for different values of hidden states (N)

Affective State	Best Parameter Values
High Interest	N=8 & T=32, N=9 & T=64, N=10 & T=64
Interest	N=8 & T=32 N=9 & T=24,88 N=10 & T=64
Low Interest	N=9 & T=32,64,88 N=10 & T=24,32 N=11 & T=64
Taking A Break	N=9 & T=32 N=10 & T=64 N=11 & T=32,64

Table 4-5 Summarization of the generalization error results; using different values of T and the best three values of N for each affective state model

According to the results given in table 4-5, the HMM's are able to well classify sequences that range from 24 to 88 posture observations; it means, the models start to differentiate the sequences after accumulating posture observations for at least 3 seconds. It suggests that with a sampling under 3 seconds it is not possible for the HMM to capture the posture patterns correlated with the established affective states.

After testing several combinations of each set of model parameters, we found the optimal combination is sampling with $T = 64$ and having a value of N=9 for the high interest, N=11 for low interest, and N=11 for the taking a break models. In particular, for the interest model, although we tried several combinations of *N* and *T*, none of them had good performance. Tables 4-6 to 4-9 give the confusion matrices of the 4 top performers.

	Classification				
Data Set	**HI**	**I**	**LI**	**TAB**	**% Recognition**
HI	117	21	20	14	68.02%
I	399	406	255	313	29.57%
LI	30	43	105	59	44.30%
TAB	30	28	27	91	51.70%

Table 4-6 Confusion matrix and performance results for a model with T=32 and N=8 for High Interest (HI), N=8 for Interest (I), N=9 for Low Interest (LI), and N=9 for Taking a Break (TAB)

	Classification				
Data Set	**HI**	**I**	**LI**	**TAB**	**% Recognition**
HI	74	10	10	7	73.27
I	196	268	149	120	36.56
LI	57	20	31	28	22.79
TAB	9	15	11	69	66.35

Table 4-7 Confusion matrix and performance results for a model with T=64 and N=9 for High Interest (HI), N=10 for Interest (I), N=9 for Low Interest (LI), and N=10 for Taking a Break (TAB)

	Classification				
Data Set	**HI**	**I**	**LI**	**TAB**	**% Recognition**
HI	42	34	7	18	41.58%
I	150	162	292	129	22.10
LI	12	16	97	11	71.32
TAB	1	14	8	81	77.88%

Table 4-8 Confusion matrix and performance results for a model with T=64 and N=10 for High Interest (HI), N=10 for Interest (I), N=11 for Low Interest (LI), and N=11 for Taking a Break (TAB)

	Classification				
Data Set	**HI**	**I**	**LI**	**TAB**	**% Recognition**
HI	77	15	5	4	76.24%
I	190	126	291	126	17.19%
LI	9	15	103	9	75.74%
TAB	2	12	8	82	78.85%

Table 4-9 Confusion matrix and performance results for a model with T=64 and N=9 for High Interest (HI), N=10 for Interest (I), N=11 for Low Interest (LI), and N=11 for Taking a Break (TAB)

Specifically, the sequences from the interest class were most of the time confused between high interest and low interest classes. This result suggests the sequences coming from the interest class are a mix of the other two; for this reason, we decided not include interest class data for further analysis and only evaluate the sequences coming from the classes of high interest, low interest and taking a break.

Using the model parameters described above, the log likelihood for an observed posture sequence is computed using the forward-backward procedure [7] on each HMM. We compare the log likelihood to label the sequence as one of the three classes.

4.4.2. Evaluation and Results

The system was evaluated as follows: First, taking all the posture sequences coming from 8 subjects that were selected randomly, and then using k-fold cross-validation with k equals 8, the system recognition accuracy was tested. Specifically, the data were divided in 8 groups, and subsequently, the HMM's were trained using the data from 8 groups but reserving one group for testing. This was repeated for all the 8 groups in the data set. Second, the classifiers were tested using posture sequences coming from two subjects that were not in the training set (new subjects).

According to the first evaluation, the system could recognize the posture sequences corresponding to the *high interest* class with an accuracy of 85.39%. For posture sequences belonging to the *low interest* class the accuracy was 74.55%, whereas an accuracy of 86.81% was obtained for posture sequences corresponding to the class of *taking a break*. Table 4-10 presents details of the recognition results for this evaluation.

	Classification			
Data Set	**HI**	**LI**	**TAB**	**% Recognition**
HI	76	6	7	85.39%
LI	22	82	6	74.55%
TAB	8	4	79	86.81%
			Total	82.25%

Table 4-10 Recognition results obtained from testing the system with data coming from 8 subjects using k-fold cross validation. High Interest (HI), Low Interest (LI), Taking a Break (TAB)

Regarding the second evaluation, the posture sequences corresponding to the *high interest* class were recognized with an accuracy of 83.33%, just 2.06% less than the first evaluation result. For posture sequences belonging to the *taking a break* class the recognition accuracy was 76.92%, whereas for posture sequences corresponding to the *low interest* class the recognition dropped to 69.23%. Table 4-11 presents the recognition results for the second evaluation. These results show the overall recognition accuracy was 76.49%; it dropped 5.76 % when the system was tested with the new two subjects.

	Classification			
Data Set	HI	LI	TAB	**% Recognition**
HI	10	1	1	83.33%
LI	5	18	3	69.23%
TAB	1	2	10	76.92%
			Total	**76.49%**

Table 4-11 Recognition results obtained from testing the system with two new subjects. High Interest (HI), Low Interest (LI), Taking a Break (TAB)

4.6 Summary

This chapter has explained the system for the automated recognition of postures in real time. Also, it has analyzed -over time- postural behaviors when a single child is working on a math-based problem in front of the computer.

In particular, the overall system exposed in this chapter is divided in three main parts: (1) the extraction of features coming from two matrices of pressure sensors mounted on the seat and back of a chair; (2) the classification of static postures using a feed-forward neural network; (3) the analysis over time of posture sequences associated with some of the affective states found by the teachers in the study developed in chapter 3 of this thesis. The results obtained for each of these parts are summarized as follow:

1. Data features were extracted by modeling each of the two-pressure matrices with the Gaussian Mixture algorithm [43] fixed with four gaussians.

2. A recognizer based on a 3-layer feed-forward neural network [44] was build. This recognizer takes as an input each of the four gaussian parameters (prior probability, mean, and variance) extracted from the two matrices of pressure sensors. It classifies the input data determining the static posture in real time, and it achieves an overall recognition of 87.64% when it is tested on children that it has not seen in the training data.

3. Using a Hidden Markov Model for representing each affective state, detectable dynamic posture patterns were found for the classes of *high interest*, *low interest* and *taking a break*. In specific, the dynamic system could recognize with an overall accuracy of 82.25% new posture sequences coming from subjects with

who the system was trained. Whereas, an overall recognition accuracy of 76.49% was obtained when it was tested with posture sequences coming from two new children that were not included in the training set.

Finally, since the viewpoint of pattern recognition, this result is particularly relevant because it is basis on a natural data set, which makes the problem much harder than only using a data set without prompt and unexpected movements.

Chapter 5

Conclusions and Future Work

5.1 Conclusions

This thesis has investigated the relationship between patterns of postural behaviors and affective states, focusing on those behaviors associated with interest and boredom that can be sensed by a chair when a child is in a computer-learning situation. The primary contribution of this thesis is the finding of different patterns of behavior for high interest, low interest, and taking-a-break, and the development of a new machine analysis algorithm for the automated detection of these different posture patterns.

For eliciting natural occurring behaviors during a learning-computer task, an experiment with children between 8 and 11 years old was conducted. In this experiment, 10 children were engaged to play for approximately 20 minutes a constraint satisfaction computer game, which we had previously determined had a high probability of eliciting the affective states of interest and boredom. In particular, for preserving as much as possible the original behavior, children were not aware until later that the purpose of the experiment was the study of their postures. Each child's session was videotaped and the postures were captured using two matrices of pressure sensors mounted on the seat and back of the chair on which each child was sitting.

Using video data captured during the children's experiment, two studies were carried out; one of them was for establishing the set of basic postures and the other one for determining the affective states to be used by the system. Specifically, in the first study, 2 human subjects –without any particular background- labeled the children's postures. In the second one, 3 expert teachers labeled the children's affective states during the learning-computer interaction.

From the first study it was found that nine postures were frequently repeated during the experiment. Hence, a posture recognition system that distinguishes this set of nine postures was built. This system achieves an overall accuracy of 87.64% when tested with children's postures that were not included in the training set. This result is significant considering that it was obtained using a data set containing the natural occurring postures gathered during the experiment, which we believe makes the problem more difficult than using a data set without such fast and unexpected movements. This posture recognition system runs in real-time, and it has been proved to work in a user-independent way. It is currently trained on children and not on adults, but potentially the same algorithms could be used to re-train the system for any population of interest.

In the second study, it was found that the three teachers could reliably recognize the states of high interest, interest, low interest, taking a break and boredom. Even though the affective state of boredom was reliable identified –every one agreed when a child was bored-, teachers only labeled very few episodes of this state. In contrast, they consistently identified an increased frequency of the taking a break state and longer periods of low interest states around fragments where they said a child was bored. However, it is important to highlight that it does not mean the taking a break state is always an indicator of boredom. For example, when the game was finished and after the child had been working hard for long time, teachers considered that it was necessary for him to take a break. Thus, the results above suggest that the boredom state seems to be a meta-class of the other two states.

This thesis has never assumed that postures can reliably reveal what a student is feeling inside. Rather, the patterns observed in the dynamics between changes of the student's postures were found to reveal significant information related to some affective states.

This thesis examined the dynamics of the ten students' posture sequences that were captured during the experiment. Specifically, the posture sequences were analyzed using a Hidden Markov Model (HMM) for representing each of the affective states identified reliably by the teachers. From this analysis, it was found that the classes of high interest, low interest, and taking a break were classified by the computer with high accuracy, while the class of boredom could not be reliable identified (the labeled boredom sequences were so few, which made it impossible to train the HMM adequately). Most of the time when the child was attending to the task, the teachers labeled the child's state not as high interest or as low interest, but just as "interest." For these segments, the computer classification was also poor; in short, the computer performed best at finding deviations from this typical state: recognizing behaviors indicating high interest, low interest, and taking a break.

The recognition results for the dynamic system were 85% for high interest, 75% for low interest, and 87% for taking a break – an overall of 87%, when the dynamic system was tested with new posture sequences coming from students that were included in the training set. The recognition results were 83% for high interest, 69% for low interest, and

77% for taking a break, - an overall of 76%, when the system was tested with posture sequences coming from two subjects that were not included in the training set These results can be compared with those from the experiment where teachers were assessing the children's affective states with an overall agreement of 79%.

As result, it seems to be that the system had a reliable classification from posture patterns of at least the states of high interest and taking a break. Thus, in contexts where children are learning while using computers, this system can provide substantial information about whether the computer is truly engaging the child or whether the frequency of taking a break is increasing. These two states may be particularly relevant for determining when not to interrupt, or when it is likely that the child might welcome an interruption. With future work, we expect that the combination of these results with other modalities (face, computer task behavior, and possible conversational input) will further disambiguate the child's state, and improve the ability of the computer to respond in a way that facilitates a productive and enjoyable learning experience for the child.

5.2 Future Work

The framework developed in this thesis open many questions as well as immediate directions for future work.

Having proposed a new approach for extracting the features coming from the two matrices of pressure sensors, the first future work direction is to benchmark this algorithm with some other implementations. Although, it was found that the new approach performed better recognizing new children's postures than just using PCA algorithm, this new approach was not compared with the most recent algorithm proposed by Slivosky and Tan [20] (see section 2.3).

In particular, in doing the comparison it is important to consider that the recognition results reported from both approaches were by using databases that have several differences. First, the matrices of pressure sensors were gathered using different kinds of chairs, one using a Steelcase Leap chair [56] that has a firm seat-pan -this thesis- and the other one using a Herman Miller Aero chair [55] that has a soft seat-pan. Second, as was discussed before, this thesis used a database that contains naturally gathered children's postures, whereas the other is based on adult postures made on command – and thus, relatively posed.

A second future work goal is testing the algorithm when the number of sensing points decreases. In particular, it can be useful for exploring the possibility of developing a less expensive version of this sensor.

Another area for future explanation regards exploring potential improvements to the classification algorithm. Since the system uses a neural network for classifying the features coming from the feature extractor, it gives to the HMM just one final posture class. Hence, it could be interesting to explore how the HMM performs when it takes the probability distributions of the posture classes.

Regarding the study for coding the affective states, teachers were labeling with considerable agreement fragments of video where all of them said a child had "other" affective state. Hence, another suggestion for future work could be the analysis of the "other" affective state classes that teachers found and investigating whether it is correlated with measurable patterns of children's behaviors.

In addition, it may be possible to build a classifier for "boredom" by combining the classifiers for low-interest and taking a break, both of which were detected with significant rates by the machine analysis. The difficulty, however, is in obtaining accurate labels of the true boredom state, especially since it is a state that seems to be socially unacceptable to show, and one which teachers tend to be reticent to identify, preferring to label multiple events of "taking a break" and "low interest" before eventually using the label "boredom."

It is also worth noting that the Hidden Markov Models used in this thesis for analysis can also be used to generate sequences of postures consistent with an affective state; thus, they can also be used for the synthesis of postures that a synthetic agent might sequence through when acting highly interested, etc. Exploring the results of this thesis for posture synthesis is another area of possible future research.

Finally, this thesis presents an analysis and development of a system that can infer from posture significant information about the child's affective states. It is relevant to emphasize that this system was tested alone, without having a multi-sensor framework. Hence, its performance when a computer agent uses it for making more complex interpretations about a child's learning experience remains to be evaluated.

Bibliography

[1] Ekman, P. & Friesen, W. (1974). *Detecting deception from the body and face*, Journal of Personality and social psychology, 29, pp. 288-298.

[2] Babad E. (1990). *Measuring and changing teacher's differential behavior as perception of students and teachers*, Journal of Educational Psychology, 82, pp. 683-690.

[3] Bull P. E. (1984). *The communication of emotion*, Paper presented at the annual conference of the British Psychological Society, University of Warwick.

[4] Smith, J. (2000). *GrandChair: Conversational Collection of Family Stories*, M.S. thesis, Media Arts and Sciences, Massachusetts Institute of Technology, Media Laboratory, Cambridge, MA, USA

[5] Kort, B., Reilly, R., Picard, R. W. (2001) *An Affective Model of Interplay Between Emotions and Learning: Reengineering Educational Pedagogy-Building a Learning Companion*, Proceedings of IEEE International Conference on Advanced Learning Technologies, Madison, USA.

[6] Tekscan (1997). *Tekscan Body Pressure Measurement System User's Manual*, Tekscan Inc., South Boston, MA, USA.

[7] Rabiner, L. R. (1989). *A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition*, Proceedings of the IEEE, 77, no. 2, pp. 257-285.

[8] Kapoor, A., Mota, S., & Picard, R. W.(2001) *Towards a Learning Companion that Recognizes Affect*, AAAI Fall Symposium 2001, North Falmouth, MA, USA.

[9] Ekman, P. & Friesen, W. (1969). *Nonverbal leakage and clues to deception*, Psychiatry, 32, pp. 88-106.

[10] Babad, E., Bernieri, F., Rosenthal, R. (1989). *Nonverbal communication and leakage in the behavior of biased and unbiased teachers*, Journal of Personality and Social Psychology, 56, pp. 89-94.

[11] Allen, V.L, & Atkinson, M. L. (1981). *Identification of spontaneous and deliberate behavior*, Journal of Nonverbal Behavior, 5, pp. 224-237

[12] Ekman, P. (1984). *Expression and the nature of emotion*, K. Scherer & P. Ekman (Eds.), Aproches to Emotion (pp. 319-343). Hillsdale, NJ: Lawrence Erlbaum Associates.

[13] Ekman, P., Friesen, W. V. (1971). *Constants across cultures in the face and emotion*, Journal of Personality and Social Psychology, 17, pp. 124-129.

[14] Ekman, P. (1965). *The differential communication of affect by head and body cues*, Journal of Personality and Social Psychology, 2, pp. 726-735.

[15] Bull P.E. (1983). *Body movement and interpersonal communication*, Chichester: John Wiley & Sons Ltd.

[16] Mehrabian A., & Friar J. T. (1969). *Encoding of attitude by a seated communicator via posture and position cues*, Journal of Consulting and Clinical Psychology, vol. 5.

[17] C. Rich, R. C. Waters, Y. Schabes, W. T. Freeman, M. C. Torrance, A. R. Golding, and M. Roth, *An animated online community with artificial agents and spoken interaction*, IEEE MultiMedia, Vol. 1, No. 4, Dec. 1994. MERL-TR94-06.

[18] Cassell, J., Nakano, Y., Bickmore, T., Sidner, C., Rich, C., *Non-verbal Cues for Discourse Structure*, Association for Computational Linguistics Joint EACL - 2001 ACL Conference.

[19] Tan H. Z., Ifung Lu and Pentland A. (1997). *The chair as a novel haptic user interface*, In Proceedings of the Workshop on Perceptual User Interfaces, Banff, Alberta, Canada.

[20] Hong Z. Tan, Slivovsky Lynne A., and Pentland Alex. (2001). *A sensing Chair Using Pressure Distribution Sensors*, IEEE/ASME Transactions on Mechatronics, vol.6, no. 3.

[21] Woolfolk, A., & Galloway, C. (1985). *Nonverbal Communication and the study of teaching*, Theory Into Practice, pp. 77-84.

[22] Doyle W. (1977). *Paradigms for teacher education research*, In L. Shulman (Ed.), Review of research in education, 5, pp. 163-198.

[23] Berliner, D. (1983). *Developing concepts of classroom environments: some light on the T in classroom studies of ATI*, Educational Psychologist, 18, pp. 1-13.

[24] Doyle, W. (1983). *Academic Work*, Review of Educational Research, 53, pp. 159-200.

[25] Goleman, D. (1995). *Emotional Intelligence*, Bantam Books, New York.

[26] Zajonc, R. B. (1998). *Emotions*, Gilbert, D. T., Fiske, S. T., and Lindzey, G. *(Eds.)* The Handbook of Social Psychology, 4th Edition, 1, pp. 591-632.

[27] LeDoux, J. E. (1994). *Emotion, memory and the brain*, Scientific American, June 1994, pp. 50-57.

[28] Damasio, A. R. (1994). *Descartes' Error: Emotion,* Reason and the Human Brain, G. P. Putnam's Sons, New York.

[29] IBM Blue Eyes Camera. http://www.almaden.ibm.com/cs/blueeyes

[30] Edmark, Fripple Place. http://www.riverdeep.net/edconnect/sofwareactivities/critical_thinking/fripple_place.html

[31] Suen, Hoi K., Ary Donald (1989). *Analyzing quantitative behavioral observation data.* Hillsdale, NJ: L. Erlbaum, Gutman Education.

[32] Kagan, J. (1994). *The Nature of the Child: Tenth anniversary edition,* New York: Basic Books.

[33] Agamanolis Stefan, MIT Media Lab (2001). URL: http://isis.www.media.mit.edu/

[34] Bakeman, R., Gottman, J. (1986). *Observing Interaction: An Introduction to Sequential Analysis*, Cambridge University Press.

[35] Mahrabian, A. (1968). *Inference of attitude from the posture, orientation and distance of a communicator*, Journal of Consulting and Clinical Psychology, pp. 293-308.

[36] Frey, S., Jorn, U., & Daw, W. (1980). *A systematic description and analysis of nonverbal interaction between doctors and patients in a psychiatric interview*, Corson, S.A. (Ed.): Ethology and Nonverbal Communication in Mental Health. New York: Pergamon Press, pp. 231-258.

[37] Birdwhistell, R. L. (1971). *Kinesics and Context*, London: Allen Lane.

[38] Bull E. P. (1987) Posture and Gesture, vol.16, Pergamon Press.

[39] A. Wilson and A. B. (1995). *Learning Visual Behavior for gesture analysis*, Proceedings of the IEEE International symposium of Computer Vision, Coral Gables, FL, USA, November 1995.

[40] Peter J. B. Hancock, A. Mike Bruton, and Vicky Bruce. *Face processing: Human perception and principal component analysis.* Memory and Cognition, Vol. 24, pp. 26-40.

[41] Helander, M.G., Paul, R.D., Morrow, J. and Zhang, L. (1997). *Comfort and discomfort at work*, The Nurturing and Pampering Paradigm, Proceedings of 13th Congress of the International Ergonomics Association. Tampere, Finland.

[42] Serra, J. (1982). *Image Analysis and Mathematical Morphology*, Academic Press, London.

[43] Richard A. Redner and Homer F. Walker.(1984). *Mixture densities, maximun likelihood and the EM algorithm*, SIAM Review, 26, no.2: 1995-239, April 1984.

[44] Bishop, Christopher M. (1995). *Neural networks for pattern recognition*, Oxford: Clarendon Press; New York: Oxford University Press.

[45] MacKay, D. J. C., (1992). *Bayesian interpolation*, Neural Computation, 4, no. 3, pp.415-447.

[46] Hagan, M. T., Menhaj, M. (1994). *Training feed-forward networks with the Marquardt algorithm*, IEEE Transactions on Neural Networks, 5, no.6, pp. 989–993.

[47] Hagan, M. T., Demuth, H. B. and Beale, M. H. (1996). *Neural Network Design*, Boston, MA: PWS Publishing.

[48] Robson C. (1993). *Real Word Research: A Resource for Social Scientist and Practitioner Researchers*, Oxford: Blackwell.

[49] Drake, A.W. (1967). *Discrete-state Markov Processes*, Chapter 5 in Fundamental of Applied Probability Theory. New York, NY: McGraw-Hill.

[50] Jaakkola, T., Meila, M., and Jebara, T. (1999). *Maximum entropy discrimination*, Neural Information Processing Systems, vol.12, (NIPS'99).

[51] Baum, L. E., Sell, G. R.(1968). *Growth functions for transformations on manifolds*, Pac. J. Math., 27, no.2, pp.211-227.

[52] Hjorth. J.S.U. (1994). *Computer Intensive Statistical Methods Validation, Model Selection, and Boot-strap.* London: Chapman & Hall.

[53] Jeker, J., Maccoby, N., & Breitrose, H. S. (1964). *Teachers accuracy in assessing cognitive feedback from students,* Journal of Applied Psychology, 48, pp. 393-397.

[54] Cassell, J., Nakano, Y., Bickmore, T., Sidner, C., and Rich, C. (2001). *Annotating and Generating Posture from Discourse Structure in Embodied Conversational Agents,* Workshop on Representation, Annotating, and Evaluating Non-Verbal and Verbal Communicative Acts to Achieve Contextual Embodied Agents, Autonomous Agents 2001 Conference, Montreal, Quebec.

[55] Herman Miller Aeron Chair: http://www.hermanmiller.com

[56] SteelCase Leap Chair: http://www.steelcase.com

[57] Evreinov G., Agranovski A., Yashkin A., Evreinova T.(1999). *PadGraph,* Proceedings of HCI International '99,vol. 2, pp. 985-989, Munich,Germany

[58] Howard Demuth & Mark Beale, Matlab Neural Networks Toolbox version 4.0, *Neural Network Toolbox User's Guide,* COPYRIGHT 1992 - 2001 by The MathWorks Inc., http:\\www.mathworks.com

[59] Plutowski, M., Sakata, S., White, H. (1994). *Cross-validation estimates IMSE.* J.D. Cowan,G. Tesauro, & J. Alspector (editors), Advances in Neural Information Processing Systems 6, pages 391-398. Morgan Kaufman.

[60] Kevin Murphy's Matlab, *Hidden Markov Model Toolbox*, http://www.cs.berkeley.edu/~murphyk/Bayes/hmm.html

[61] Sierra Entertainment, *The Incredible Machine*, http://www.incrediblemachine.net

[62] Goldin-Meadow, S., D. Wein, et al. (1992). *Assessing Knowledge Through Gesture: Using Children's Hands to Read Their Minds.* Cognition and Instruction **9**(3): 201-219.

[63] Goldin-Meadow, S., M. W. Alibali, et al. (1993). *Transitions in Concept Acquisition: Using the Hands to Read the Mind.* Psychological Review **100**(2): 279-297.

Appendix A: Game

The measures from the game that were considered are described below. I use these measures as they characterize the child's performance and the individual differences.

1. The *Status Indicators* (see table 3.12) were taken as the game independent variables. Each of these variables were assigned with a constant integer for being used to compute the overall game score. Tables A-1 to A-2 show the game independent variables and their values.

Game Independent Variables	**Value**
Failure	-3
Hint Error	-2
Hint Help	-1
New Game	0
Game Running	1
Level of Difficulty Low	2
Level of Difficulty Medium-Low	4
Level of Difficulty Medium-High	6
Level of Difficulty High	8
Success	10

Table A-1 Game Independent Variables and their Constant Values

STATUS	DESCRIPTION
New Game	When a new game starts, which could be either the beginning of the interaction; start of a new game, or just after a change in the level of the difficulty.
Game Running	When the child is solving a game. This state excludes the events when the child asks for hints or when the child is checking its solution.
Failure	When the child checks the solution by pressing the "Check it" button and the solution is incorrect. The first time the button is pressed and the solution is incorrect, then the program simply tells that the solution is incorrect. If the child makes some changes and checks the response for the second time, and it is wrong again, then the program shows all the mismatched Fripples.
Success	When the child succeeds in solving the game.
Hint Help	When in the middle of the game the "Hint/Check It" button is pressed, the program tells the number of the mismatched Fripples.
Hint Error	When right after Hint Help state the "Hint/Check It" button is pressed again, the program shows one of the mismatched Fripples.
Level of Difficulty	When the level of difficulty is changed. There are four levels of difficulty: low, medium-low, medium-high, and high. The level of difficulty can be changed through a window that appears when the level of difficulty button is pressed. Each time the level of difficulty is changed, the old game is abandoned and a new game starts.

Table A-2 The Fripple Place game events description

2. The *game score* is considered a game-dependent variable and is calculated using the status indicators that occur during the whole interaction. The aim of computing the score is to have a measure of the child's game performance. Each child starts with a score equal to zero and as the game evolves and the independent variables (status indicators) appear,

the score changes. Table A-3 shows the number of points assigned when each status variable occurs.

Status Indicator	**Num. Of Points Assigned to Score**
New Game	0
Game Running	1
Success	Successes * Level of Difficulty
Hint Help	-1
Hint Error	-2
Failure	-3

Table A-3 Number of points assigned when each status variable emerges

3. Number of Games played during the whole interaction
4. Duration of each game.
5. Number of times the child asked for hint help or hint error.
6. Number of failures.
7. Number of successes.
8. Number of times that the level of difficulty was changed.
9. Level of difficulty sequence

Appendix B: Models Parameters

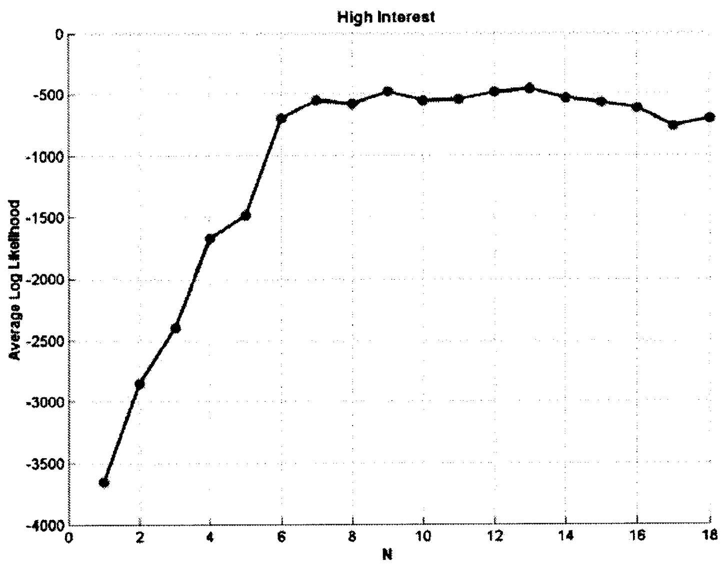


Figure B-1 *High interest* model generalization errors for different values of hidden states (N)

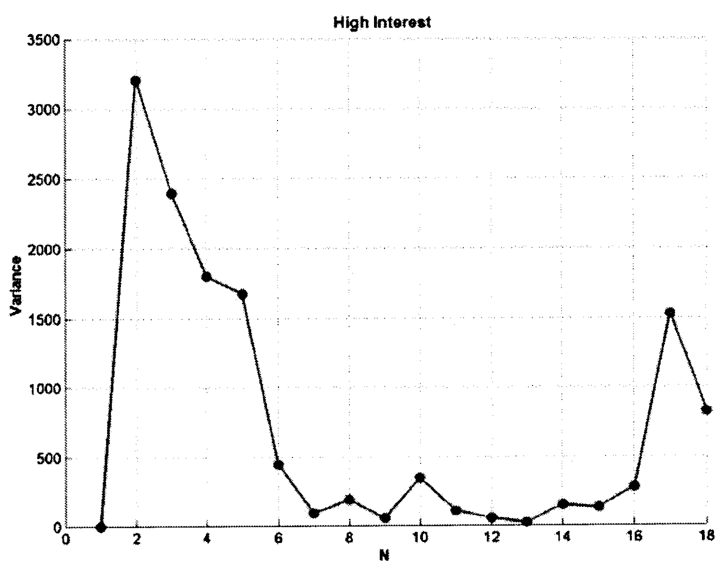


Figure B-2 Variances of the *high interest* model generalization errors for different values of hidden states (N)

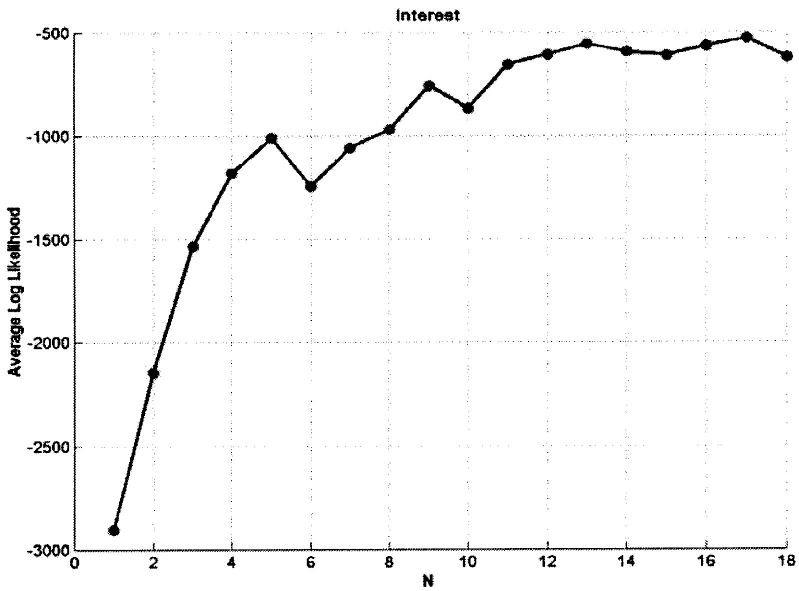


Figure B-3 *Interest* model generalization errors for different values of hidden states (*N*)

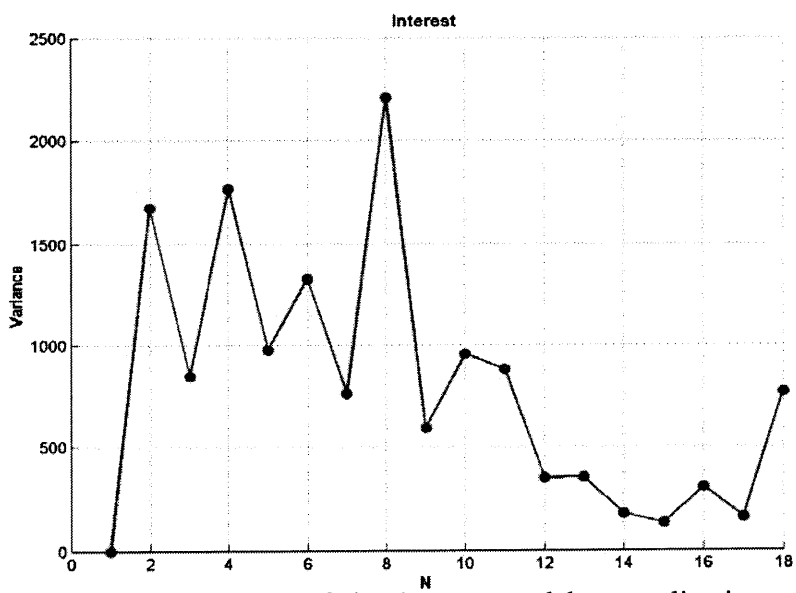


Figure B-4 Variances of the *interest* model generalization errors for different values of hidden states (*N*)

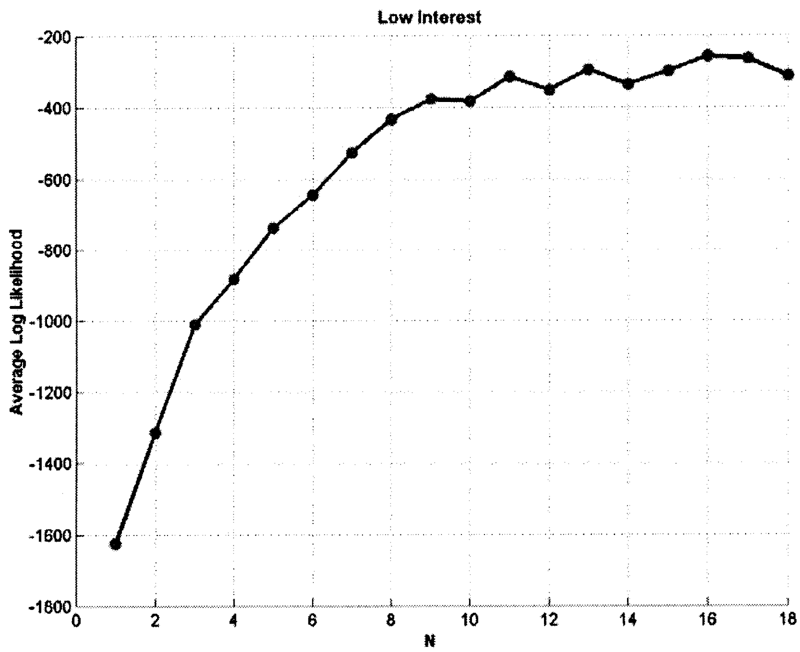


Figure B-5 *Low interest* model generalization errors for different values of hidden states (*N*)

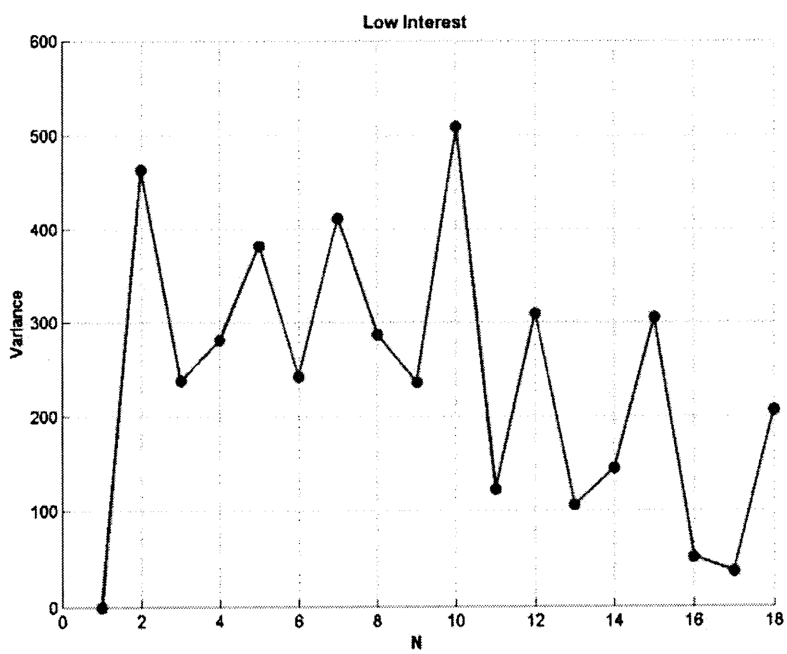


Figure B-6 Variances of the *low interest* model generalization errors for different values of hidden states (*N*)

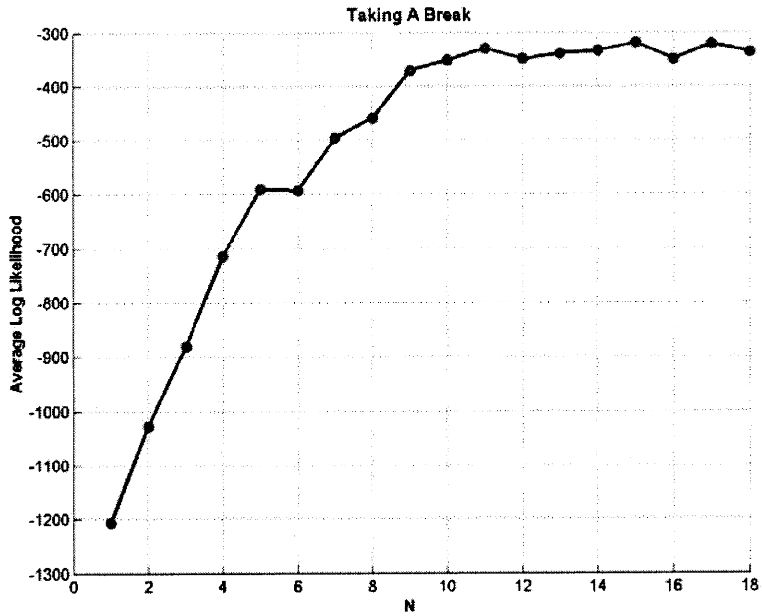


Figure B-7 *Taking a Break* model generalization errors for different values of hidden states (*N*)

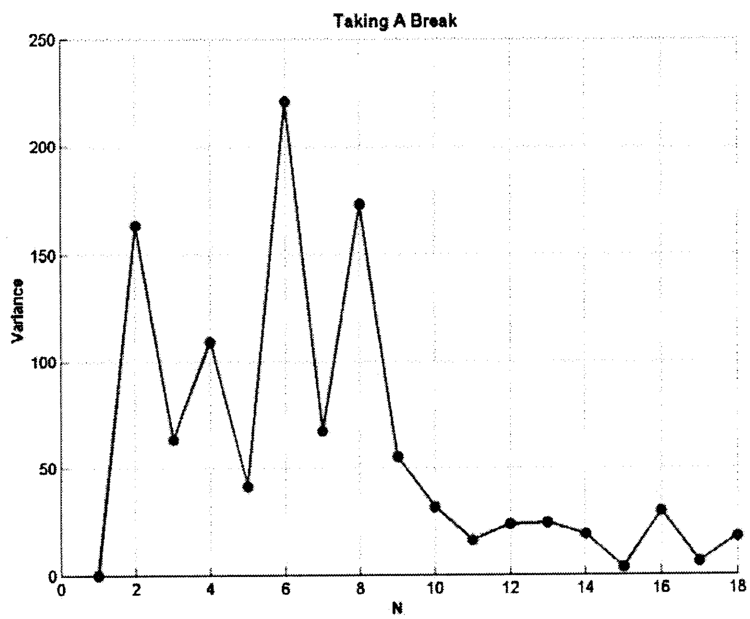


Figure B-8 Variances of the *Taking a Break* model generalization errors for different values of hidden states (*N*)

Appendix C: Cohen's Kappa

Cohen's Kappa is a statistical method that assesses inter-judges agreement for nominally coded data. It can be applied at both the global level (i.e. for the coding system as a whole) and the local level (i.e. for individual categories). In either case, the formula is

$$\text{kappa} = \frac{(\text{po - pc})}{(1\text{ - pc})} \qquad \text{Equation-6}$$

where po is the proportion of units that the two judges coded the same, and pc is the proportion expected by chance. An equivalent formula, using frequencies, is

$$\text{kappa} = \frac{(\text{fo - fc})}{(\text{N - fc})} \qquad \text{Equation-7}$$

where fo denotes the number (not proportion) of units coded similarly, fc represents number of units that would be expected to be coded the same way by chance alone, and N is the number of units coded by either coder (i.e., if they code 50 units each, N= 50, not 100). In this thesis the method based on frequencies was used.