A Novel Approach for Attention Management in E-learning Systems

Gennaro Costagliola, Mattia De Rosa, Vittorio Fuccella
Department of Mathematics and Informatics
University of Salerno
{gencos,vfuccella}@unisa.it

Nicola Capuano, Pierluigi Ritrovato
DIIMA - University of Salerno
{ncapuano,pritrova}@unisa.it

Abstract

This study presents new approaches for the detection and treatment of the attention of a student by an e-learning system through the use of the information given by the implicit interaction of the student with the system and the data coming from non-invasive devices such as webcams.

Furthermore, the paper proposes two models for the treatment of the attention of students to be applied to an existing e-learning environment, in order to provide personalized content to the students and thus improving their learning experience.

1. Introduction

Over the past ten years e-learning has evolved from early systems to *Intelligent Tutoring Systems* (ITS), *Smart Classrooms* and *Mobile Learning* (e-Learning with mobile devices). Today, e-learning aims to be strongly student-centered, in order to provide a personalized learning experience. Its principal objectives are then not only to foster successful learning but also to involve students in the learning process and maximize their interest [15].

An expert teacher easily captures the emotional state of students and adapts lessons accordingly, in order to maximize their interest and participation. An e-learning system in order to provide a quality educational experience should be able to behave similarly [2].

This need is reflected in the development of systems capable of detecting the attention of a student during e-learning sessions. The use of biometric sensors can provide good information on the student's emotional state [7], but at the same time it can create physical discomfort to the student in addition to possible hardware costs and logistic problems. It is therefore advisable to use non-invasive systems such as log analysis [6] and cheap hardware already present on computers [18] such as a webcam [3, 4].

This document briefly describes the e-learning system *IWT* (*Intelligent Web Teacher*) [9, 5, 14], an ITS that provides personalized content to the student, and *WiSe*, a system that detects the attention of a student through the analysis of the IWT interaction log and of the webcam video of the student during the e-learning sessions.

Finally we present a basic and an extended model for the integration of the student attention detection module with the e-leaning environment, in order to customize the learning content and improve the learning experience.

The contribution of the research presented in this paper lies in the definition of novel models for the application of automatic detection of student attention to e-learning intelligent tutoring systems, and in particular to the IWT system.

This paper is organized as follows. Section 2 discusses the state of the art of attention detection and its use in e-learning environments. Sections 3 and 4 briefly describe the characteristics of *IWT* and *WiSe*. Sections 5 and 6 present the two alternative models for managing the attention in the IWT e-learning environment. Finally, section 7 offers our conclusions and describes further work.

2. Related Work

In literature there are several studies dealing with the detection of the emotional state of computer system users. Only some of them focus on the detection of the user's attention.

In [8] the authors postulated the existence of a direct link between the user's comfort and emotional state. A research team at Purdue University has instead developed a system for the perception of posture [17, 19] through the analysis of the pressure exerted by the body on the chair. For this purpose they used a chair properly equipped with pressure sensors. The system performs real-time analysis using computer vision techniques to map pressure distribution obtained from sensors located on the seat and backrest.

An interesting conceptual study in [11] theorizes that some macro movements on the chair are indicators of the emotion and therefore suggests the possibility of creating an office chair that can adapt itself to the user's emotional state.

The Affective Computing Group of the MIT Media Labs, has carried out an interesting project concerning the perception of the state of interest (high, medium, low, bored, taking a break) in children analyzing their sitting position

[12]. Starting from the results of [17], they exploit a neural network for classification of posture and a Hidden Markov Model for recognizing the state of interest. Through the use of the pressure maps they claim they can obtain an accuracy of 82% for known subjects and 76% for unknown subjects.

In [10] the authors deal with the creation of computing and communication systems that can detect and reason on the human attention by fusing the information received from multiple sources. A probabilistic model combines data from sensors, from user interaction with the system, from the previous pattern of activity and attention in order to estimate the user's attention and then to adapt the system behavior.

In [3] the authors present a neuro-fuzzy approach to infer the attention level of a user in front of a monitor using a simple camera.

In [4] the authors estimate the level of attention/interest of a user who reads the text on the computer using a camera to detect the position and movement of certain points around the eyes and the position of the iris. The system analyzes user behavior and provides a model of six general learners' states (Frustrated/Struggling to read, Tired/Sleepy, Not paying attention, Distracted, Attentive, Full of interest). The authors present a case study where an e-learning system provides modifications to the presentation of the text according to the level of attention detected in children with dyslexia.

In [13], the authors describe the use of physiological signals to improve student interactions with character-based interfaces that adapt themselves to reflect the user's affective state. The paper presents "Emotion Mirror", an example of a system where emotions are sent back to the user and its evolution "Emphatic Companion", an agent (represented by a character) that adapts its behavior according to the emotional state of the user, e.g. giving support and encouragement.

As a part of a learning content recommendation system, the work presented in [15] uses biometric sensors to detect the emotional state of the student in order to adjust the content offered by the system. By comparing sessions implementing emotion detections to sessions without detections, they see that the manual interventions of the student (required when the system does not automatically provide the content that the student needs) are reduced by 91% in the first case.

From the literature we can note that even though there exist studies that treat the automatic detection of attentional/emotional states in e-learning, however they do not show how to apply it to existing e-learning intelligent tutoring systems.

3. Background: The IWT Platform

IWT, [9, 5, 14], is an e-learning intelligent tutoring system whose aim is to customize the learning experience to the real needs and preferences of the student. The innovative features of IWT compared to other e-learning solutions can be summarized as follows:

- possibility of automatic or assisted generation of learning paths from the learning objectives;
- ability of automatically customize courses based on previous knowledge of the individual students and their learning preferences;
- possibility of content management at a high level of abstraction using ontologies;

IWT models the knowledge through Learning Objects (*LOs*), Metadata and Ontologies. The LOs are the basic teaching modules that can be used during learning. The Metadata formally describe the LOs through a standard set of attributes. In particular, IWT adopts the IEEE LOM [1] standard that specifies the description of the LO through 47 items grouped into 9 categories. Ontologies represent the teaching domains and offer knowledge management concept-oriented support at a higher abstraction level.

IWT is able to (automatically) capture the learning preferences and student acquired knowledge during their educational experience. IWT manages this informations using three elements: a Cognitive State, that represents the knowledge possessed by the student by concept-vote pairs, a set of Learning Preferences that refer to fields of the Educational category defined by the IEEE LOM metadata and a set of Evolution Rules used for updating the student's cognitive state at the end of each verification test. In particular, the Educational category fields correspond, among the others, to *Interactivity Type*, *Learning Resource Type*, *Interactivity Level*, *Semantic Density*.

In IWT a course is generated by a set of Objective Concepts defined by the teacher or by the student him/herself. It first builds the best learning path for a given student considering his/her Cognitive State (eliminating already known concepts and adding any missing pre-requisites). From the learning path so constructed, it generates the best presentation for a given student by considering his/her Learning Preferences and choosing, therefore, the LO more congenial to him/her.

In the course, in general, the fruition begins with the first LO and continues until it reaches a verification test (Milestone). When the test ends, adjustments are made to the learning path portion not yet viewed by the student to respond to any weaknesses identified through the insertion of recovery LOs.

4. WiSe attention detection system

In order to obtain information about the student attention, IWT uses the WiSe platform services. The communication between IWT and WiSe follows the request-reply model (IWT requests the student attention when it needs and WiSe replies with the attention level detected in that moment).

WiSe monitors students during learning sessions and estimates their attention level by considering:

- the capture of half-length figure of the student with a webcam (video tracking);
- the capture of the computer screen and of keyboard and mouse generated input (input tracking);
- the capture of information about the tasks simultaneously active on the user computer (activity tracking).

Basing on these measurements, WiSe calculates a current attention level that can be one of the following 4 discrete values: high, medium, low, distracted. This value is calculated by combining the results of three types of analysis on available measurements: posture analysis, implicit interaction analysis and concurrent activities analysis.

The fusion is performed through a statistical model. Details about the three types of analysis are given below.

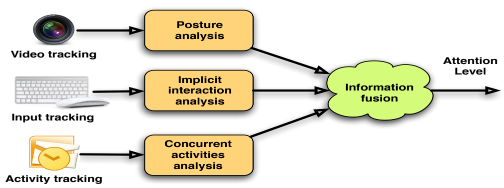

Figure 1. Elaboration process for the calculation of the attention level

4.1. Posture analysis

WiSe determines a first approximation of the attention level through the detection and the analysis of the user sitting posture in front of the monitor. The posture is deduced by analyzing the sequence of images taken by a frontal webcam through head tracking algorithms capable of detecting the position and the orientation of the head in six degrees of freedom.

Through classification algorithms, values representing the position and the orientation of the head are mapped on 9 major and 9 minor poses. Information about the gaze direction (to the screen or elsewhere) is added to the detected pose in order to infer the visual attention focus. The obtained information, in accordance with [12], is analyzed using pattern recognition techniques able to extract an indication of the perceived attention level.

4.2. Implicit interaction analysis

Implicit interactions between the user and the system are a useful indicator of the user's attentional state. WiSe, in particular, measures and processes the following parameters: the TSR (Time Spent for Reading), the TSS (Time Spent for Scrolling), the link clicking/link following time, the downtime, the time spent moving the mouse, the number of clicks, the page resize events and the keyboard typing.

WiSe is able to track and process these parameters, applying a statistical model that relates the implicit interaction events with the attention level. In the elaboration process, the collected data are related to the context information reflecting the specific content received, the environmental conditions and the characteristics of the user obtained from the student model.

4.3. Concurrent activities analysis

For the measurement of attention WiSe also uses an evaluation of the load required by tasks simultaneously active on the user computer. In fact it is widely acknowledged that the voluntary attention, involving cognitive processes, has a finite capacity. The WiSe analysis currently focuses on only two types of events: messages exchanged by e-mail and instant messaging tools.

Through automatic text analysis techniques, WiSe evaluates the marginal significance of information compared to the cost of the interruption to determine whether it is useful information or noise. In particular, incoming messages are classified by measuring the relevance of their content with the content of active LOs and the membership of senders to the same educational context of the user (e.g. teachers, tutors, fellow students).

5. The attention management: the basic model

Depending on the level of attention estimated by WiSe the IWT e-learning platform performs actions aimed at improving student learning.

This section describes our basic model for the management of attention by IWT. In this model, given the current level of attention (as received by WiSe) and the latest recorded levels of attention, IWT decides whether to present the student with a set of seven possible actions in order to raise his/her attention. This model does not require to change the WiSe attention detection system and allows the student to choose how to proceed. This prevents IWT from producing "wrong" actions. Furthermore, since alerts may distract/bother the student (as indicated in [10]), the model tends to minimize its occurrences.

5.1. The attention analysis

In this section we show how to calculate two indices, the *weighted average attention* (waa) and the *attention trend* (at) to represent the attention profile of the students during their learning session. Both of them are calculated taking into account the attention levels returned by WiSe every 20 seconds. This time interval has been defined by analyzing the data of the IWT learning sessions of 37 students. In fact, we observed that the attention used to change every 43.8 seconds on average with a standard deviation of 41.7. Furthermore, we noted that the total time of the attention states with a duration greater than 20 seconds accounted for the 92% of the total duration of the sessions.

The indices are calculated by taking into account the last 15 attention levels and by assigning a numerical value to each possible attention level, as follows:

high	average	low	distracted
1	0.67	0.33	0

For each of the 15 observations (from the oldest attention level to the newest) we assign an increasing weight calculated according to the function $f(x) = \frac{x}{n}$ ($n = \sum_{i=1}^{15} i = 120$), as shown in the following:

Observation no.	1	2	3	4	5
Weight	0.008	0.017	0.025	0.033	0.042
Observation no.	6	7	8	9	10
Weight	0.050	0.058	0.067	0.075	0.083
Observation no.	11	12	13	14	15
Weight	0.092	0.100	0.108	0.117	0.125

Note that this weight assignment gives more importance to the latest observations.

We then calculate the first index by the formula $waa = \frac{\sum_{i=1}^{15} att_i \times weight_i}{k}$ where $k = \sum_{i=1}^{15} weight_i$, att_i is the attention value in the i-th observation, and $weight_i$ is the weight assigned to the i-th observation in the table. Since $k = 1$, we have $waa = \sum_{i=1}^{15} att_i \times weight_i$. The left chart in Figure 2 shows an example with some sample data.

In order to calculate the attention trend, for each observation, we calculate the difference between the value of the attention and the weighted average attention and then multiply it by the observation weight, that is $a_i = (att_i - waa) \times weight_i$.

Finally, on the values (i, a_i), we first draw the line corresponding to the linear regression (see the right chart in Figure 2) and then we calculate the difference between the y values of the line for $i = 15$ and for $i = 1$. It is possible to prove that this difference ranges between approximately -0.08 and $+0.08$ where a positive value indicates a growing trend of attention, while a negative value a decreasing trend.

In order to have discrete values for the two indices we define two thresholds $0 < t1 < t2 < 1$ for the weighted average attention and two more thresholds $-0.08 < r1 < 0 < r2 < +0.08$ for the attention trend. The thresholds allow us to define the discrete values as shown in the following:

waa	$0 \leq waa < t_1$	$t_1 \leq waa < t_2$	$t_2 \leq waa < 1$
Value:	low	medium	high

at	$-0.08 < at < r_1$	$r_1 \leq at \leq r_2$	$r_2 < at < +0.08$
Value:	decreasing	stable	increasing

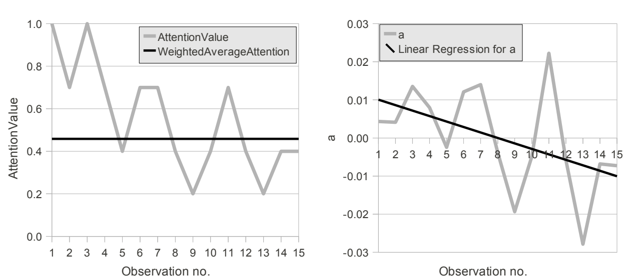

Figure 2. Weighted average of the attention and the attention trend on sample data

5.2. The use of the Attention Profile

Once IWT has calculated the student attention profile it needs to decide how to react in order to improve the student learning experience. In this model IWT will have to decide whether to present the student with a set of possible actions, without being too annoying. The decision is based on the two following conditions:

- the pair *weighted average attention*/*attention trend* has value medium/decreasing or low/stable or low/decreasing, and
- at least two minutes have passed since IWT has presented the list of actions for the last time.

In the case the student has lately chosen not to do any action then the minutes of waiting will be doubled.

These choices are motivated by the fact that if the attention trend is increasing we do not need to operate, if it is stable we need to operate a change only if the attention is low, and if it is decreasing then we do not need to operate only when the attention is high. Please note that the system will act only in three cases out of nine and only when the student has not been recently helped, and this is done in order to be as less invasive as possible.

In the following we give the list of the IWT possible actions to be selected by the inattentive student:

1. calculate a new learning path to start after the current LO, using different parameters from the original one (the student indicates if the current LO type of interaction is pleasant/unpleasant, if the level of interactivity is too low or too high, if the LO is too easy/difficult, if the semantic density is too high or too low; according to these data, the system obtains new values for the interaction type, interaction level, difficulty level, semantic density);
2. show a LO on the same subject but with different metadata (the student chooses from a list showing the other LOs available and their interactivity type, interactivity level, difficulty, semantic density);
3. insert a test in the learning path after the current LO;
4. display a LO on a correlated topic selected by the system based on its ontology:
5. provide a moment of relax/pause at the end of the current LO;
6. change the LO presentation mode (font size, etc.);
7. do nothing (the student can also indicate whether he was really distracted and why).

The actions will be presented to the student in the form of a simple questionnaire.

6. The attention management: the extended model

As in the basic model the e-learning platform IWT, depending on the level of attention estimated by WiSe, performs actions aimed at improving student learning.

This section describes our extended model for the management of the attention by IWT. Differently from the basic model, in this extension IWT does not always present the student with a list of actions, but it has to perform most of the actions (in order to raise his/her attention) without user interaction, according to the present and past levels of the student's attention. This avoids the risk of distract/bother the student.

This model also requires an extension of the WiSe attention detection system in order not only to discover the attention level but also to detect its cause. This gives more information when choosing the next action to take. The extension requires that the attention state is represented by a pair of values:

- the attention level: a real value between 0 (low) and 1 (high);
- reason of attention: a justification value for an attention level (bored, distracted, sleepy, frustrated (too difficult topic), search on Internet for information related to the course, chat about topics of the course, perform actions not related to the current topics, disturbed by someone in the classroom, idle, absent).

Note that this model involves the risk of performing "wrong" actions, risk that IWT decreases using a statistical analysis on students' previous interactions.

6.1. The attention analysis

As in the basic model IWT requests the attention value every 20 seconds. The attention profile includes the *weighted average attention*, the *attention trend* (calculated similarly to the basic model) and the most relevant *reason of attention*. We also calculate and store the attention trend after the execution of each action, in order to provide useful information to IWT to choose the next actions.

6.2. The use of the Attention Profile

Once IWT has calculated the student attention profile it needs to decide how to react in order to improve the student learning experience. In this model IWT will have to decide whether to perform actions and, if so, which ones by taking into account the logged data from previous IWT decisions.

After an initial training phase aimed at initializing the log, IWT executes the algorithm in figure 3 to calculate the *weighted average attention trend* for each action. The algorithm takes as input the current LO, the current student, the current attention profile and the set of log entries defined by the tuple (*executed action*, *student*, *LO*, *attention profile before the action*, *attention trend after the action*). Moreover it makes use of the two functions *attentionProfileSimilarity* and *LOSimilarity*. Both functions return a value between 0 and 1 but while LOSimilarity calculates the similarity between the metadata of two LOs, attentionProfileSimilarity calculates the similarity between two attention profiles.

for all action a from $actionList$ **do**
  $i \Leftarrow 1;$
  for all log entry le where $le.executedAction = a$ and $le.student = currentStudent$ **do**
    $w[i] \Leftarrow$ (LOSimilarity$(currentLO, le.LO)+$ attentionProfileSimilarity$(currentAttentionProfile, le.attentionProfileBeforeTheAction))/2;$
    $at[i] \Leftarrow le.attentionTrendAfterTheAction;$
    $i \Leftarrow i + 1;$
  end for
  $av1 \Leftarrow \frac{\sum_{n=1}^{i-1} w[n] \times at[n]}{\sum_{n=1}^{i-1} w[i]};$
  $i \Leftarrow 1;$
  for all log entry le where $le.executedAction = a$ and $le.LO = currentLO$ **do**
    $w[i] \Leftarrow$ attentionProfileSimilarity$(currentAttentionProfile, le.attentionProfileBeforeTheAction);$
    $at[i] \Leftarrow le.attentionTrendAfterTheAction;$
    $i \Leftarrow i + 1;$
  end for
  $av2 \Leftarrow \frac{\sum_{n=1}^{i-1} w[n] \times at[n]}{\sum_{n=1}^{i-1} w[i]};$
  $weightedAverageAttentionTrend[a] \Leftarrow \frac{av1+av2}{2};$
end for

Figure 3. Algorithm that calculates the past attention trend for each possible action

The algorithm takes into account for each type of action:

- the logs entries related to the current student;
- the logs entries related to the current LO.

In the former case the weighted average of the logged attention trend (*av1*) after an action execution is calculated using as weights the similarity between the current LO and the LO indicated in the log and the similarity between the current attention profile and the attention profile stored in the log. In the latter case *av2* is calculated using as weights only the similarity between the current attention profile and the attention profile stored in the log.

The first value provides information on the attention trend observed in the past for the current student with similar attention profile/LO, while the second provide informations on the attention trend observed in the past for other students with similar attention profile on the current LO.

IWT then executes the action with the highest calculated weighted average attention trend if at least two minutes have passed since the latest executed action. It can be noted that the system dynamically adapts itself to the student needs and to the current LO in order to maximize the future attention trend.

In this extended model the possible actions are both selected and executed by IWT. The list of actions are as follows:

- calculate a new learning path to start after the current LO, using different parameters from the original one (e.g. different values for interaction type, interaction level, difficulty level, semantic density);
- insert a test in the learning path after the current LO (e.g. for doubtful cases);
- provide a moment of relaxation/pause at the end of the current LO (e.g. if too tired);
- show an alert message to call a distracted user;
- show an alert message that asks the user whether s/he wants to take a LO on another topic selected by the system (e.g. when the user uses a search engine to search for something on the basic topics of the LO);
- show an alert message that asks the user whether s/he wants to see a LO with different metadata on the same topic instead of the current LO (selectable from a list);
- change the LO presentation mode (font size, etc.);
- do nothing.

7. Conclusions and future research

In this study we described the IWT e-learning environment, an ITS that can customize the learning content according to the profile of the student. We then described the WiSe system for the attention detection of the IWT students inferred from the information granted by the implicit interaction of the student with the system and data coming from non-invasive devices (webcam).

Finally we presented two alternative models for the management of the attention by IWT, in order to dynamically adapt the content presented to the students thus improving the learning experience.

Future research will aim at a more formal validation of the presented models, by organizing e-learning sessions of tests on IWT/WiSe with students from the University of Salerno.

8. Acknowledgement

This work was realized within the WiSe "Wireless Services: an User Centric Approach" project co-financed by the MIUR (D.Lgs. N. 297/1999).

This research has been partially supported by the grant "cofinanziamento per attrezzature scientifiche e di supporto, grandi e medie (2005)" from the University of Salerno.

References

[1] Ieee standard for learning technology-extensible markup language (xml) schema definition language binding for learning object metadata. *IEEE Std 1484.12.3-2005*, pages 0_1–46, 2005.

[2] S. Alexander, A. Sarrafzadeh, and C. Fan. Pay attention! the computer is watching: Affective tutoring systems. In A. Rossett, editor, *Proceedings of World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education 2003*, pages 1463–1466, Phoenix, Arizona, USA, 2003. AACE.

[3] S. Asteriadis, K. Karpouzis, and S. Kollias. A neuro-fuzzy approach to user attention recognition. In *ICANN '08: Proceedings of the 18th international conference on Artificial Neural Networks, Part I*, pages 927–936, Berlin, Heidelberg, 2008. Springer-Verlag.

[4] S. Asteriadis, P. Tzouveli, K. Karpouzis, and S. Kollias. Estimation of behavioral user state based on eye gaze and head pose–application in an e-learning environment. *Multimedia Tools Appl.*, 41(3):469–493, 2009.

[5] N. Capuano, M. Gaeta, A. Marengo, S. Miranda, F. Orciuoli, and P. Ritrovato. LIA: an intelligent advisor for e-learning. *Interactive Learning Environments*, 17(3):221–239, 2009.

[6] M. Cocea and S. Weibelzahl. Can log file analysis estimate learner's level of motivation? In *Proc. of 14th Workshop on Adaptivity and User Modeling in Interactive*, pages 32–35, 2006.

[7] C. R. C. Conati and H. Mclaren. A study of using biometric sensors for monitoring user emotions in educational games. In *Proc. of Workshop on Modeling User Affect and Actions: Why, When and How*, 2003.

[8] P. A. Fenety, C. Putnam, and J. M. Walker. In-chair movement: validity, reliability and implications for measuring sitting discomfort. *Applied Ergonomics*, 31(4):383–393, 2000.

[9] M. Gaeta, F. Orciuoli, and P. Ritrovato. Advanced ontology management system for personalised e-Learning. *Knowledge-Based Systems*, 22(4):292–301, 2009.

[10] E. Horvitz, C. Kadie, T. Paek, and D. Hovel. Models of attention in computing and communication: from principles to applications. *Commun. ACM*, 46(3):52–59, 2003.

[11] P. V. F. K. C. K. J. Overbeeke. The emotion-aware office chair. In *Proc. International Conference on Affective Human Factors Design, Asean Academic Press, London*, 2001.

[12] S. Mota and R. W. Picard. Automated posture analysis for detecting learner's interest level. In *Computer Vision and Pattern Recognition Workshop, 2003. CVPRW '03. Conference on*, volume 5, pages 49–49, 16-22 2003.

[13] H. Prendinger, J. Mori, S. Mayer, M. ISHIZUKA, and R. Member. Character-based interfaces adapting to users' autonomic nervous system activity. In *Proceedings of the Joint Agent Workshop (JAWS-03)*, 2003.

[14] E. Sangineto, N. Capuano, M. Gaeta, and A. Micarelli. Adaptive course generation through learning styles representation. *Universal Access in the Information Society*, 7(1):1–23, 2008.

[15] L. W. M. S. R. Shen. Affective e-learning: using "emotional" data to improve learning in pervasive learning environment. *Educational Technology Society*, 12-2:176–189, 2009.

[16] H. Tan, L. Slivovsky, and A. Pentland. A sensing chair using pressure distribution sensors. *Mechatronics, IEEE/ASME Transactions on*, 6(3):261–268, sep 2001.

[17] S. P. Tarzia, R. P. Dick, P. A. Dinda, and G. Memik. Sonar-based measurement of user presence and attention. In *Ubicomp '09: Proceedings of the 11th international conference on Ubiquitous computing*, pages 89–92, New York, NY, USA, 2009. ACM.

[18] M. Zhu, A. M. Martinez, and H. Z. Tan. Template-based recognition of static sitting postures. In *Computer Vision and Pattern Recognition Workshop, 2003. CVPRW '03. Conference on*, volume 5, pages 50–50, 16-22 2003.