On detecting user's attention from physical activities for attention-based interfaces

By

Amina Hassan Abubakar

Department of Computer Science,
Faculty of Physical Sciences,
Ahmadu Bello University, Zaria.
Email: ahabubakar@abu.edu.ng

ABSTRACT

Modeling user's attention state can immensely help in the development of attentive device (s) that can proactively present user with correct information when and where needed. Detecting how user attends to devices and interfaces is a crucial problem in user interface design; especially for ubiquitous systems. To help inform the design of such systems, it is important to know how a user pays attention to an interface. Gaze was used in several researches to determine user's attention, but we argue that gaze is not always practical, especially if the communication is non-visual. In this paper, a method that detects and infers user's attention from activity information obtained from body-worn sensors was used. Three different experiments were conducted to investigate the effect of varying audio signals to user's attention, and a classification system based on Gaussian naïve bayes method was implemented in MATLAB for detecting change in user activity pattern. The correlation between such changes and the occurrence of an audio signal was shown. The System was evaluated, and an overall accuracy of 80.04% from experiments 1 and 2, and 79.36% for experiment3 was obtained. Based on the result and the accuracy of the classifier in detecting user's points of attention, it was concluded that body-worn sensors could be used to detect attention to audio interface from user's activity; hence physical activity can be used for Attention-based interfaces.

Keywords: Attention, Attentive User Interface, Body-worn Sensors, Gaussian Naïve Bayes classifier

INTRODUCTION

Emergence of ubiquitous computing has necessitated a new method of interacting with computing devices, which resulted in moving from the explicit method of input through the keyboards and other pointing devices to a more implicit method of input which could be provided as a result of our natural interactions with devices and our environment.

Recent developments in miniaturization of computing devices over the last decade aimed at achieving device ubiquity, has resulted in an increased interest in researches on ways of making computing devices understand where the user's focus of attention is at any given time, the type and amount of work he is performing, the work expected completion time and the person's level of interruptability. Because despite the fact that we are witnessing device ubiquity, these devices are still not properly equipped to negotiate communication with the user Vertegaal, Shell, Chen and Mamuji (2006) which is becoming a big issue especially with

device interconnectivity, with so many of them interrupting user, seeking for user's attention without considering the availability of user. As observed in Garlan, Siewiorek, Smailagic, and Steenkiste, (2002) human attention is a limited resource that must be properly utilized, hence to enable devices to proactively decide best time to interrupt the user; a complex assessment of the information content and the context needs to be carried out.

Detecting and inferring attention is very important in the design and development of attentive applications and technologies. Actions and uttered words can provide sufficient information for communicating one's intention – which forms the basis of most the work done in context-aware applications today. Action of attention which can be sensed in several ways that include one's location, movement, tasks, utterances etc.; can provide a very rich clue to one's intention. Thus, our actions can reveal excellent clues to what we are attending to, which makes detecting attention from human physical activities feasible.

Gaze was used in (Maglio, 2000; Garlan, Siewiorek, Smailagic, and Steenkiste, 2002; Mamuji, Vertegaal Shell, Pham, and Sohn, 2003; Vertegaal, 1999) to model user's attention and / or designed attentive interface. But gaze is not always practical especially when communicating with non-visual interface, such as audio interface. With body-worn sensors rich information about user's state of attention can be obtain by monitoring changes in person activities pattern, because it is not possible for one to attend to all things at the same time, these changes can provide us with clues of where the user's attention is at any point. Three users wore the Mobile Sensing Platform (MSP) Choudhury, Borriello, Consolvo, Haehnel, Harrison, Hemingway, Hightower, Klasnja, Koscher, LaMarca, Landay., LeGrand, Lester, Rahimi, Rea, and Wyatt, (2008) of which one is male and two females. The female subjects wore the MSP on their head scarves and the male subjects attached it to the hat.

The aim of this research is to determine user's attention from physical activities; which can be achieved through the following objectives:

i. To establish the feasibility of using user's activity to infer attention through the use of body-worn sensors.
ii. Implementing a classification algorithm for detecting the correlation between the occurrence of audio signal and change in user activity pattern.
iii. Investigating with different scenarios the effect of varying audio signals to user's attention.
iv. Evaluate the performance of the classifier.

The following hypotheses were formulated from the stated objectives:

i. Patterns of body movement are affected by auditory events.
ii. Change in body movement can be an indication of acknowledgment of auditory event.

Review of related work

There is an increased interest in researches that detect, and or model attention especially in Ubiquitous Computing (UbiComp) area of computer scienceAccording to Vertegaal (2003) maximizing productivity through the use of

computers was hindered by cost incurred as a result of to the ill equipment of today's user interface to meet up with current challenges of ubiquitous computing which happens as a result of them not being change for more than two decades. Making devices aware of their context of use will certainly narrow the gap between them and humans. According to Hinckley, Pierce, Horvitz and Sinclair, (2005), paying attention to devices while performing other real-world activities such as washing, cooking, walking along a street, running, word processing, browsing etc., which demand attention themselves may over burden to a person.

Ward, Lukowitz, and Troster, (2005) used 3-axis accelerometer and a Microphone to recognize assembly plant activities such as screw driving, drilling, grinding etc. Wearable reader called iGlove and RFID chip were used for activity recognition. They used Triaxial Accelerometers for recognizing activity using three classifiers known as Decision Tree (DT), Multi-layer perception (MLP), and Support Vector Machine (SVM).
In a research carried out by Vertegaal (1999) to investigate attention in a multi-party communication using a Gaze hardware setup comprising of Eyegaze computer, speaker, Microphone snapshot Camera, Infrared Camera and a prototype system "The Gaze Groupware System", he showed how conveying gaze direction can solve the problem of multi-party communication: establishing who is speaking or listening to whom. Gerasimov and Bender (2000) used sound to detect attention, and they conclude that creative sounds in device-to-human communication can improve its usability. In the work of Sallen (1995) which involve a setup of multiple monitor/speaker /camera, the Hydra System was described in which head orientation, relative position and gaze are preserved at the time of multiparty videoconferencing. Selker (2004) designed an eye-base attentive interface "Eye-aRe" that uses gaze to infer attention and looking away as indicating lack of interest (attention).

The work of Duffner, and Garcia (2016) and Recasens (2015) used face images to infer coarse gaze directions. The advantage of the method is that it relies on a state-of-the-art appearance-based gaze estimator to obtain the initial features for the unsupervised gaze target discovery. To evaluate the benefits of this approach, for this baseline we directly used the face features f extracted from the CNN model as input to the clustering.

Raw gaze direction has been used to estimate visual attention on public displays in (Sugano, Zhang, and Bulling, 2016; and Huang, Kwok, Ngai, Chan, and Leong, 2016). The physical size of the target object and its position related to the camera was manually measured and projected the object as bounding box on the camera image plane. The input image was then classified as eye contact if the estimated gaze location was inside the bounding box. Their method assumed accurate knowledge of the target object location.

Zhang, Sugano, and Bulling, (2017) presented a method for eye contact detection that combines a state-of-the-art appearance-based gaze estimator with an unsupervised gaze target discovery approach. Their method was evaluated in two real-world scenarios: detecting eye contact at the workplace, which include the main work display, cameras mounted to target objects, as well as during every day social interactions with the wearer of a head-

mounted egocentric camera. The performance of the method was empirically evaluating in both scenarios and demonstrate its effectiveness for detecting eye contact independent of type and size of the target object, camera position, user and recording environment.

METHODOLOGY

The method used in this research involve designed of a model of attention that detect attention points of from activity using the example of communication involving many people in a cocktail party, in which a person is able to focus person attention to only one person's speech at a time. We applied this approach to augment user's attentive capability. In this work, we used to turn of the body or part of it during activity to provide us with information on what the user is attending to.

The first hypothesis we are interested in testing is finding out whether patterns of body movement are affected by auditory events or not. That is, we want to know if the user is doing some activities say for example person is in a quite office environment with little or no possibility of distractions performing some task (such as web browsing, spreadsheet or word processing etc.) and suddenly another person tries to communicates to the person, possibly by calling the person's name, what does the person do? Does person turn the entire body or part of the body; may be the head? Or if a user is walking and the mobile phone rang, or there is a message notification, does person turn to where the phone is in order to pick it up and answer the call or to read the message? If person did turn, can we detect that with body worn sensors?

Our second hypothesis has to do with finding out if we did detect change in body movement can it be an indication of acknowledgment of auditory event, can we find the correlation between such changes in body movement and the occurrence of an auditory event?

To achieve the four high level aims identified in section one, series of tasks were carried out some of which were described in this section, and some in later sections even though some of them were carried out in parallel.

Three different experiments were conducted on three subjects where two different scenarios were investigated. In the first two experiments, user with the MSP attached to the head scarf sat in a quite lab performing some data processing operations, with the person's face and body facing the Laptop directly. Then the controller of the experiment made some sounds at random of varying type and pitch aimed at distracting the user's attention from the person's current activity (normal activity), each time taking the snapshot of the timestamp. For each of three experiments, three sets of data were collected each of which lasted for 3-4 minutes.

In the third experiment the user is also wearing the MSP as in the first and second, but for this experiment, user was walking along the corridor, the controller also made different attempts to distract user by calling the person's mobile phone, name from different directions, and making other sounds like hitting other objects etc.

Attention Labeling Technique

In order to observe human attention from activities, it is important to label the perceive attention points which will provide

a way of validating the results obtained from the classification system later during the study. After studying the various options that could be used for labeling human attention, which we adapted from those used in labeling activities discussed in Tapia (2003) that includes: Direct / indirect observation, self-report; recall surveys, self-report; Time Diaries, and end of study interviews.

We combined both direct and indirect method of labeling attention. During the study, we observed user's tasks, behavior and focus of attention as well as how that changes as a result of distraction. In addition, we used stopwatch to record when the distractions occur. The timestamp at which the sound was created was recorded. Similarly, audio signals were recorded throughout the duration of the experiments, this also serve as a ground truth when plotted.

DESIGN AND IMPLEMENTATION

The proposed system comprises of four major components, which are: The Mobile Sensing Platform, inter-process communication application, audio recording program and attention recognition and classification algorithm

A block diagram of the attention detection and classification system is shown in figure 1 which shows the various tasks and where each was carried out within the system (that is either in the MSP or Laptop).

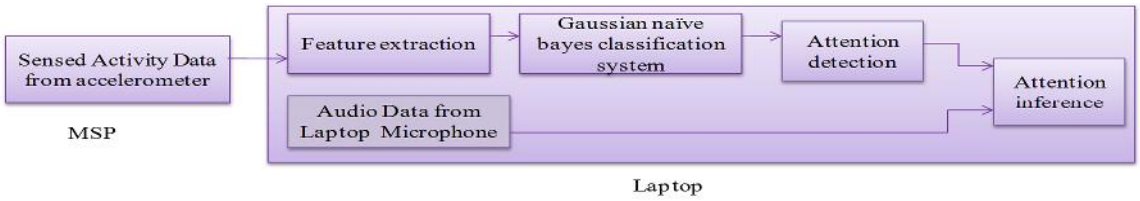

Figure 1: Block Diagram of the Attention Recognition and classification System

Attention Recognition Algorithm

The essence of using the MSP, IPC application was to obtain the necessary data needed for creating a learning algorithm that can identify points of attention from activation of sensors by user's activities. In order to achieve the high-level goals, set out in chapter one, an algorithm that correlate activity labels with sensor firing, and use such information to predict attention from new sensor firing is needed. Hence, the algorithm described in this section was developed.

Placement of sensor (MSP)

The MSP was attached to the head scarf of the female participants during the experiment, while for the male participant it was attached on their hat.

The data collected from the sensors are transferred to MATLAB for visualization and subsequent analysis. Figure 2 shows a plot of one of the raw acceleration data collected during one of experiments.

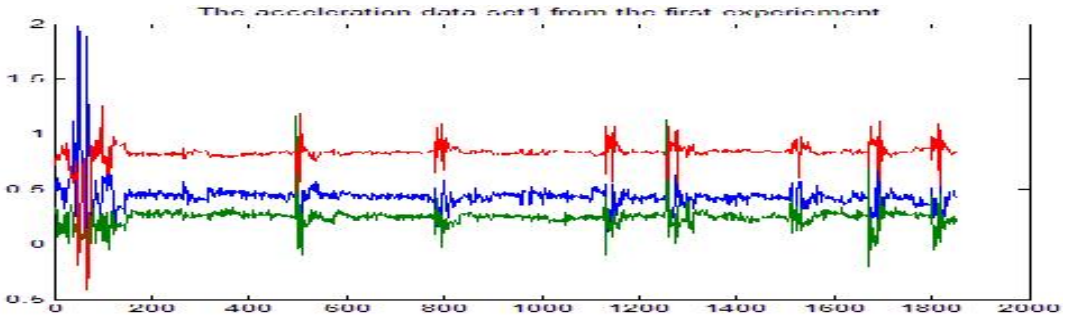

Figure 2: Plot of raw accelaration data, also showing data at synchronisation point

For all the experiments conducted, the data collected from the microphone was saved in a file in a directory of the computer. Similarly, the timestamps taken with stopwatch are recorded in text files and saved in the computer.

Off-line learning and classification

The attention classification algorithm is designed for off-line training and testing to minimize the tradeoff between feature, model and computational complexity.

Feature Extraction

During the preprocessing stage, we extracted features from the datasets that were used for training. This is to ensure that the inclusion of redundant data is minimized, that is, only relevant data is used. To achieved that, each of the data sets obtained during the experiments was plotted and its pattern examined, then different points representing the data attributes as interesting (indicating points when user turn as a result of the distraction) and not interesting (points where user is just doing normal activity) were selected and annotated. The selected interested points were all cut out from the original data set and merged together and then saved in a text file. Similarly, the not interested points were also cut out from the original data set, merged and saved in another text file. This process was repeated for all the data sets.

Implementation

A Naïve Bayes classifier for Gaussian estimators of probability density function was designed and implemented in MATLAB R2009b for recognizing human attention from activities; it uses the mean and standard deviation parameters explained above. Similarly, MATLAB was also used to display activation of the sensor, attention labels based on microphone and timestamps data acquired, the process of obtaining the training data explained above was also generated in MATLAB to test the accuracy and robustness of the classifier.

Two models; one and two for "turning activity" and "normal activity" were created that can efficiently learn from the training dataset with no additional computational cost. For each of the models the classification function explained earlier in this section is called, the testing data, some respective parameters obtained from section 4.2 and window size were passed to it. The classier assign attention when the outcome of model one is less than that of model two.

The performance of the classifier was validated on several real datasets obtained from our experiments.

RESULTS

Figure 3 shows Plot of sound data set1 from the first experinment and detected points of attention from data set1.

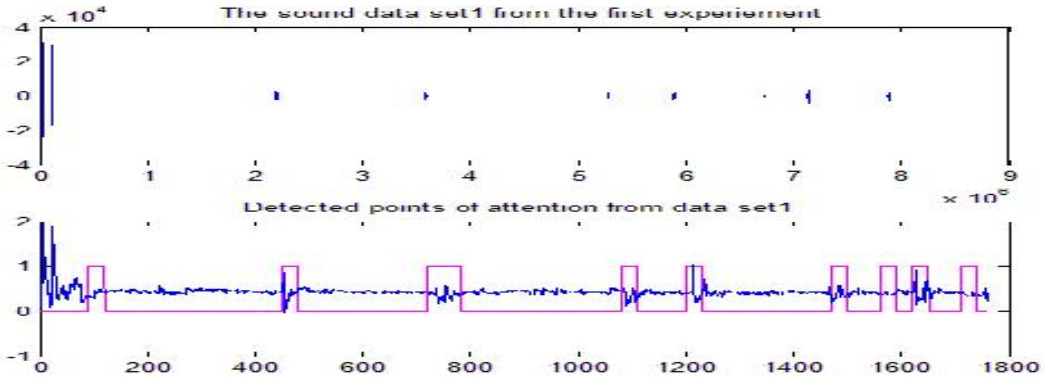

Figure 3: Plot of sound data set1 and detected points of attention

Detecting and distinguishing valid activity from false activity (such as distinguishing between true positive, false positive, true negative and false negative activity) is another challenge in activity recognition; sometimes it is possible to detect an activity when there is none or to detect no activity when there is an activity. To solve this problem, we measured the performance of our system, by calculating recall (*the measure of completeness of the detected attention points*), precision (*which is the degree of exactness of the classifier*) and the accuracy (*the degree of the result conforming to the actual value*) of the classifier in detecting when the user activity pattern changes using the formulas:

$$Recall = TP/ (TP + FN) \quad (1)$$

$$Precision = TP/ (TP + FP) \quad (2)$$

$$Accuracy = (TP + TN)/ (TP + FP + FN + TN) \quad (3)$$

Where TP means True Positive, FP means False Positive, FN means False Negative and True Negative.

Table 1: System Performance on data from Experiment 1 and 2

System Performance on data from Experiments 1 and 2				
Subjects	Experiments	Recall (100 %)	Precision (100%)	Accuracy (100%)
1	Set1	75	75	77.78
	Set2	83.33	62.5	75
	Set3	62.5	60	73.91
2	Set1	80	72.73	76.19
	Set2	85.72	85.72	91.67
	Set3	85.74	75	85.71

Result from table 1 was plotted and shown in figure 4 to illustrate the performance of the system based on data from Experiment 1 and 2.

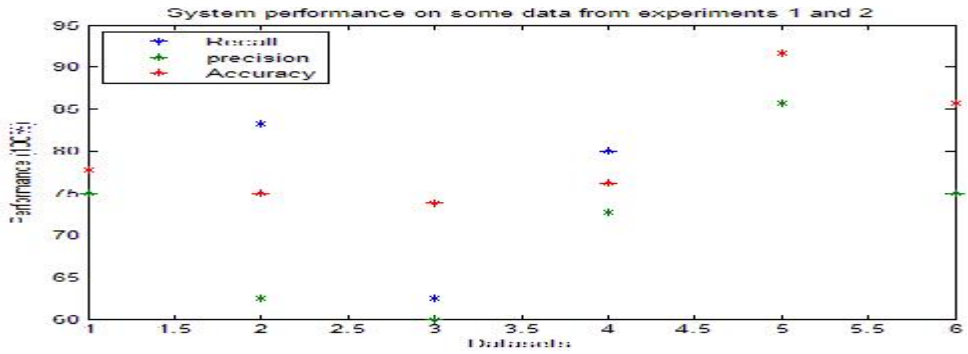

Figure 4: System performance based on data from Experiment 1 and 2.

Table 2 shows the system performance on data from Experiment 3 based on recall, precision and accuracy of the system.

Table 2: System Performance on data from Experiment 3

System Performance on data from Experiments 1 and 2				
Subjects	Experiments	Recall (100 %)	Precision (100%)	Accuracy (100%)
1	Set1	62.52	85.71	81.25
	Set2	50	60	64.28
	Set3	100	85.71	92.56

Result from table 2 was plotted and shown in Figure 5 to illustrate the performance of the system based on data from Experiment 3.

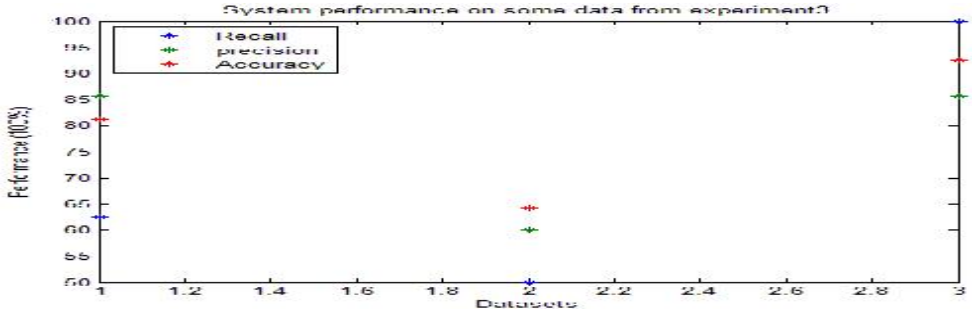

Figure 5: System performance based on data from Experiment 3.

Significance of the Result

The ability to accurately model human attention points from physical activities could enable interface designers to design attentive interfaces. Knowledge of one's attention is very helpful in designing attentive user interface that could be used in minimizing their user's cognitive workload by cancelling noise and irrelevant information before they reach the human brain.

Similarly, it can be applied in custom design, development and deployment of assistive technologies such as cognitive enhancing devices for people with short and long-time memory loss.

CONCLUSION

Change in physical activity pattern can be used in modeling attention and designing physical attentive interface, but it is noteworthy that the designed needs to be informed on how activity change is influence by attention. A model of user states of attention which infer the state of interruptability from activity was developed. This research showed that attention can also be detected from user's physical activity.

REFERENCES

Choudhury, T., Borriello, G., Consolvo, S., Haehnel, D., Harrison, B., Hemingway, B., Hightower, J., Klasnja, P., Koscher, K., LaMarca, A., Landay., J. A., LeGrand, L., Lester, J. Rahimi, A., Rea, A., and Wyatt, D. (2008) The Mobile Sensing Platform: An Embedded System for Capturing and Recognizing Activities. IEEE Pervasive Magazine – Special Issue on Activity – Based Computing.

Duffner, S., and Garcia, C. (2016) Visual focus of attention estimation with unsupervised incremental learning. *IEEE Transactions on Circuits and Systems for Video Technology*, 26 (12). Pp. 2264 – 2272.

Garlan, D. and Siewiorek, D., Smailagic, A. and Steenkiste, P. (2002) Project Aura: Towards distraction-free pervasive computing. IEEE Pervasive Computing.

Gerasimov, V. and Bender, W. (2000) Things that talk: Using Sound for device-to-device and device to human communication. IBM Systems Journal Vol 39 (3).

Hinckley, K., Pierce, J., Horvitz E. and Sinclair, M. (2005) Foreground and Background Interaction with Sensor-Enhanced Mobile Devices. ACM-TRANSACTION http://research.microsoft.com /users/kenh/papers/TochiSensing.pdf.

Huang, M. X., Kwok, T. C. K, Ngai, G., Chan, S. C. F., and Leong, H. V. (2016) Building a personalized, auto-calibrating eye tracker from user interactions. *In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, Pp 5169–5179.

Maglio P. P., Matlock T., Campbell C. S., Zhai S. and Smith B.A. (2000) Gaze and Speech in Attentive User Interfaces. Proceedings of the third International Conference on Multimodal Interfaces. Beijing China Springer-Verlag.

Mamuji, A., Vertegaal R., Shell J., Pham, T., and Sohn, C. (2003). AuraLamp: Contextual speech recognition in an eye contact sensing light appliance.

Sallen, A.J. (1995) Remote Conversations: The Effects of Mediating Talk With technology. http://research.microsoft.com/~asellen/publications/sellen.pdf

Selker, T. (2004) Visual attentive interfaces.http://pubs.media.mit.edu /bttj/Paper16 Pages146-150.pdf.

Sugano, Y., Zhang, X., and Bulling. A., (2016) Aggregaze: Collective estimation of audience attention on public displays. *In Proceedings of the 29th Annual Symposium on User Interface Software and Technology,* Pp. 821–831.

Tapia, E.M. (2003) Activity Recognition in the home Setting Using Simple and

Ubiquitous Sensors. Master of Science Thesis submitted to the Program in media Arts and Sciences, school of Architecture and Planning. Massachusetts Institute of Technology.

Vertegaal, R. (1999) The Gaze Groupware system: Mediated Joint Attention in Multiparty Communication and Collaboration. Proceedings of the SIGCHI Conference on Human Computer Interaction. dl.acm.org.

Vertegaal, R. (2003) Attentive user Interfaces. Communications of the ACM. Vol. 46 (3). Pp 30-33.

Vertegaal, R., Shell J.S, Chen D. and Mamuji A. (2006) Designing for augmented attention: Towards a framework for attentive user interfaces. Computers in Human Behavior. Attention aware systems-Special Issue: *Attention aware Systems. Vol. 22 No.4*

Ward, J.A, Lukowitz, P., Troster, G. (2005) Gesture Spotting Using Wrist-Worn Microphone and 3-Axis Accelerometer. ACM international Conference Proceeding Series; Proceeding of the 2005 joint conference on Smart object and ambient intelligent: Innovative Context-aware Services: usages and technologies

Zhang, X., Sugano, Y., and Bulling, A. (2017) Everyday Eye Contact Detection Using Unsupervised Gaze Target Discovery.

