Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Pervasive and Mobile Computing 3 (2007) 74–94
pervasive and mobile computing
www.elsevier.com/locate/pmc

Multi-modal emotive computing in a smart house environment

Simon Moncrieff*, Svetha Venkatesh, Geoff West, Stewart Greenhill

Department of Computing, Curtin University of Technology, Kent Street, Bentley 6102, Western Australia, Australia

Received 30 September 2005; received in revised form 29 June 2006; accepted 16 July 2006
Available online 11 September 2006

Abstract

We determine hazards within a smart house environment using an emotive computing framework. Representing a hazardous situation as an abnormal activity, we model normality using the concept of anxiety, using an agent based probabilistic approach. Interactions between a user and the environment are determined using multi-modal sensor data. The anxiety framework is a scalable, real-time approach that is able to incorporate data from a number of sources, or agents, and able to accommodate interleaving event sequences. In addition to using simple sensors, we introduce a method for using audio as a pervasive sensor indicating the presence of an activity. The audio data enabled the detection of activity when interactions between a user and a monitored device didn't occur, successfully preventing false hazardous situations from being detected. We present results for a number of activity sequences, both normal and abnormal.

Keywords: Assisted living; Audio; Surveillance and monitoring; Sensor network; Hazard detection

1. Introduction

This paper is concerned with the development of smart environments for the assisted living of elderly people. A particular aspect of smart environments relevant to the care of

* Corresponding author. Tel.: +61 8 9266 7680; fax: +61 8 9266 2819.
E-mail addresses: simonm@cs.curtin.edu.au (S. Moncrieff), svetha@cs.curtin.edu.au (S. Venkatesh), geoff@cs.curtin.edu.au (G. West), stewartg@cs.curtin.edu.au (S. Greenhill).

1574-1192/$ - see front matter
doi:10.1016/j.pmcj.2006.07.003

the elderly is the determination of hazards within the environment in real time. Typical hazards are appliances being left unattended, such as unattended electrical devices, or abnormal environmental states such as a front door being left open and unattended. It is important to note that estimation of a hazard must be made in the context of the user's normal behaviour, and it is this requirement that makes the problem difficult.

Most systems for activity monitoring are interested in event detection after the fact, for example every couple of hours [1]. Although this is valid in many cases, real time response in needed for many hazards at home e.g. leaving the bath to run over. Other approaches to such problems is to use temporal models of activity recognition with hidden Markov models [2,3] and build representations for normality from which abnormality can be inferred. However, the variability in behaviour patterns both within and across individuals makes it difficult to detect consistent patterns. Additionally the computational complexity makes it difficult to use in real time. Furthermore, when multiple as well as multi-modal sensors are used, a typical probabilistic based method of activity recognition starts to fail because the state space grows exponentially with the number of devices and states.

Potential hazards can be determined in many ways across a broad spectrum of modalities. At one end of the spectrum, we can have a fully monitored home with multiple cameras augmented by other sensors in the environment. At the other end of the spectrum, the home can be augmented with simple sensors such as pressure mats, simple microphones (for sound level sensing) and reed switches. All these sensors have modalities that have advantages and disadvantages. For example pressure pads give simple discrete on/off events usually with 100% reliability. Video processing is highly flexible and consequently can be used to detect many different aspects of activity. However the current state of the art in video processing is not reliable enough and the processing required is complex. Although microphones can be regarded as simple sensors when used as a sound level device e.g. for detecting loud noises, it can also be used to detect more complex audio events in various ways. They can also be used to localise sound events [4] but this requires complex processing. Ultimately, the challenge is the integration of most if not all of these different devices seamlessly into a framework for activity and hazard monitoring. When using audio and video sensors, one must be aware of the privacy issues raised and hence it is important to use these devices to *only* detect events and not record the video and sound for later playback.

In this paper we explore the integration of sound, and simple sensors such as pressure pads, reed switches and X10 devices into a multi-modal framework for hazard detection in assisted living. This is based on a recent model proposed in [5] where each device, with the potential to be hazardous, is represented as an agent. A measure of *anxiety* is associated with each agent, representing the potential hazard represented by the device. The proposed agent based approach is device centric and the state space is automatically factored, thus making the approach scalable and applicable in small as well as large pervasive environments. Whilst the previous work concentrated on using simple sensors such as pressure pads and reed switches, we explore a more comprehensive integration of the multi-modal sensor data, in the form of simple sensor and audio data, into the anxiety framework. We introduce a method for using the audio as a pervasive sensor using foreground sound events to determine audio activity. We argue that the use of sound and simple sensors can give the same functionality as the alternative approach of video processing but with much less processing costs and lower bandwidth. Additionally, we formalise the anxiety model such that any arbitrary sensor can be seamlessly integrated. The anxiety is formalised by deriving statistical models of the interactions of a user with devices within the environment.

This paper incorporates a very important modality, namely sound and in particular the detection of what are defined to be sound events indicative of normal activities by the occupant. There are two ways sound can be used. The first and more difficult approach is to recognise individual sounds and relate them to specific activities. The second and more general approach is to model background noise, from which foreground sounds can be detected. Background noise is defined as consisting of typical regularly occurring sounds such as the fridge being on and traffic. Events based on sound are different to those generated by pressure pads, reed switches and accelerometers, in that there is no spatial information available. In fact sound is picked up wherever the activity is occurring such as away from other sensors and devices and hence can detect activity when people are not interacting with devices but still moving around. The characteristics of sound implies that it provides a pervasive and contextual source of data from which we get many foreground events. This is in contrast to other discrete simple sensors, whose use is limited by practicality issues such as their placement, i.e. such sensors require an interaction between the user and a monitored object.

We detail a method to implement the system within a smart house environment. We close the loop between the anxiety model, sensors, and user by using a personal digital assistant (PDA) to communicate with the occupant. The PDA is further used as an agent in the system. An anxiety is associated with the PDA when an interaction between the user and the PDA is expected.

The significance of this paper lies in two key areas. Firstly, this paper details the implementation of a scalable, agent based, emotive computing framework for determining hazards in a smart house environment. Secondly, the paper describes the incorporation of multiple sources of sensor data into the anxiety framework, most notably the use of audio as a pervasive method for examining the activities of an user over a wide area. The novelty is threefold. Firstly, we introduce and formalise an interaction based probabilistic model for hazard detection within a smart house environment. The anxiety represents an emotive model that is scalable and independent of activity sequences. Secondly, we use the robust detection of foreground audio events for activity detection in assisted living. This enables the integration of event detection by audio into the anxiety framework demonstrating (a) the extended functionality of the model and (b) the fusion of multi-modal sensor information within the framework. Thirdly, we incorporate a PDA as a third form of sensor, providing interactive feedback from a user.

The layout of this paper is as follows. Section 2 discusses related work in the areas of audio surveillance and smart home environments, and outlines the relevant background information for the “anxious home” and the audio background modelling. Section 4 details audio activity detection and our multi-modal approach to the determination of anxiety. Section 5 describes the experimental process and results of the implementation of the multi-modal anxiety approach. Section 6 describes our implementation of the system in a smart house environment.

2. Background

2.1. The anxious home

Anxiety is an emotion in the human sense and recently, much work has been carried out into emotional computing [6], mainly to enable computers to communicate with humans. It is argued that decision making by humans requires emotions as well as rational thought for fast decision making. Hence modelling human emotions in computer decision making may improve the performance of computers for such activities.

The previous approach to hazard detection [5] within the smart house environment modelled the hazards within the environment by learning the patterns of interaction between a user and numerous simple sensors. This method can cope with many activities but relies on an occupant moving around and regularly interacting with devices monitored by sensors during normal activities. This was extended to include simple loud sound detection as well as wearable devices such as accelerometers and PDAs [7]. Loud noises were taken to indicate potential emergencies but no attempt was made to identify the type of noise. The accelerometers were used on the body to detect the occupant's stance and whether they had fallen. The PDA was used to obtain a response from the user for one of these events and an emergency invoked if no response was received from the occupant.

2.2. Smart house activity monitoring

There has been considerable research using simple sensor data to monitor activity. "Stove Guard"[1] has current and motion sensors that can turn off the stove after a certain time if it is on (current flowing), and if it is unattended (no motion detected). Combinations of simple sensors have been used for recognising activity in houses [8–10]. Glascock and Kutzik [1,11] use a small number of infra-red sensors for coarse activity monitoring that is mainly suited for making sure someone has taken medication, eaten etc. which only requires events to be reported at two hour intervals. Recently, a system [12] has been proposed that uses a form of anxiety that rises if there is little activity in the house, determined by the use of a number of standard sensors, such as window and door sensors. If the lack of activity is unusual, based on learned data, an alarm can be raised. The work presented in this paper differs from the above in that we are interested in the interactions of various devices enabling richer semantics to be inferred and monitored in real time enabling prompt responses to abnormal behaviour.

Considerable research has been performed in the recognition of the complex patterns of activities that occur in smart houses. Much of this research focuses on using various forms of Markov model: hidden Markov models, hierarchical hidden Markov models, abstract hidden Markov models etc. [2,3,13–15]. Models are constructed for each type of activity, with the best matching model for subsequently observed data being used to interpret the activity represented by the data. Hidden Markov models (HMM) can accommodate variations in the duration of activities, but are sensitive to changes in the order of sub-activities and events, and the interleaving of events. We present a method that models

[1] www.absoluteautomation.com/stoveguard — accessed Feb'05.

the duration between interactions with devices in the environment. The method is able to provide meaningful results when presented with differing sequences of interactions resulting from the interleaving of events and changes in the order of the events.

2.3. Audio surveillance and monitoring

Audio analysis methods for surveillance and monitoring have predominantly centred on the detection of specific audio cues, or sound events. Such methods include a tele-monitoring system for the detection of sound events such as cries for help [16,17], the classification of sound categories related to bathroom activity [18], and the detection of alarm sounds [19]. Cowling [20,21] proposed a method to determine a taxonomy for the classification of environmental sounds for the purpose of audio surveillance. Härmä [22] proposed a method for monitoring acoustic activity using supervised and unsupervised clustering. While these methods focus on the detection of specific sound events, our approach extends audio surveillance by deriving contextual information from the analysis of the audio signal.

2.4. Audio background modelling

We define background audio to be the persistent audio characteristics that dominate a portion of the signal. We then associate foreground audio with activities occurring within the environment. Background audio is determined using an on-line, adaptive Gaussian Mixture Model (GMM) to model the background audio, as detailed in [23]. This method was augmented by combining fragmented background models using entropy calculated between the distributions within the GMM, which results in a more robust determination of the background model. The foreground sound events are characterised by a difference in the characteristics of the audio accounting for the preceding audio context. This is a more robust approach in comparison with sound level sensing.

The algorithm enables the determination of sounds within the environment that differ from the background audio. The nature of the algorithm is such that the audio data, once processed, can be discarded, in order to reduce privacy concerns. Furthermore, the processing costs associated with the algorithm are low, approximately 0.013 s is required to process each second of audio data (determined using a Pentium 4 3.0 GHz processor running Windows XP).

3. The anxiety model

The main objective of this research is to model normality i.e. the normal activities of an occupant in their house. What we desire is a measure that will be below a threshold for normal activity but rises above the threshold for abnormal activities. Importantly we do not want to model abnormality directly. Abnormality, almost by definition, is not modellable because abnormal events rarely occur and would not be statistically meaningful. The essential idea proposed in this paper is that a device, when on, is in a hazardous state until it is switched off. Such devices are stoves, baths and fridges. The longer each device is left *unattended* the more hazardous it should become e.g. leaving a stove on for eight hours should be considered dangerous. We introduce the concept of *anxiety* here to represent the patterns of activity associated with a device. This anxiety measure essentially represents how worried the device is given that it is being ignored by the occupant. The more the occupant interacts with the device or is found to be nearby, the less anxious the device is. When a device is turned on, its anxiety is zero, but rises over time if it is not attended by the occupant. Eventually when it reaches some threshold, some action should be taken. An important issue is what is meant by attended. This can be modelled in two ways; (1) the device is directly interacted with (settings changed or occupant adjacent e.g. standing on a pressure pad next to the device), and (2) the device is observed from close range e.g. opening the fridge that is near the cooker means the occupant can check on the state of the stove easily or can get to the stove within a reasonable time to interact with it.

Each of these should mean the anxiety of the device reduces instantaneously by some amount and then starts to rise again. The second model can be thought of as a function of whether (1) the device is normally interacted with, and (2) how far away it is. By normally interacted with, we mean that, for example, when the stove is on, it is normal to visit the fridge regularly. By how far away the device is, we mean that interaction with near devices is more reassuring than with devices far away. For example, interacting with the fridge that is near means it is easy to observe the stove and the occupant can get to the stove quickly. Interacting with the bath would mean it would take longer to reach the stove and, normally, it would be difficult to observe the stove from the bathroom.

To further illustrate the concept, consider a breakfast scenario consisting of the following sequence of events:

- Occupant opens cupboard, takes cereal from cupboard, closes cupboard, puts cereal on table.
- Occupant opens fridge door, takes egg and milk from fridge, puts milk on table, put egg on stove to cook, closes fridge door.
- Occupant sits down at table, eats breakfast.
- Occupant gets egg from stove, turns off stove, sits down, eats egg.

There are two potentially hazardous situations here. The first is leaving the stove on to boil the egg dry and then melt the pan. The second is leaving the fridge open which would spoil the food, use electricity and possible burn out the motor. There are also interleaved activities here i.e. the stove is turned on while the fridge is open. Once the stove is turned on, its anxiety will start to rise. The fact that the occupant is nearby (by the table) should mean the anxiety should not rise as fast as, say, if the occupant leaves the kitchen. As long as the occupant checks on the stove regularly or finishes eating their cereal in time, the anxiety shouldn't reach the alarm level. When the occupant turns off the stove, the anxiety for the stove should go to zero as it is not in a hazardous state anymore. Once the occupant leaves the fridge door open, the anxiety for the fridge should rise and reach the alarm state. If the door is closed before it reaches this state then anxiety reduces to zero and no alarm is signalled.

The approach follows an agent based methodology. Each device with the potential to be a hazard is represented as an agent, with many agents representing the various potentially hazardous devices within the environment. Each agent continuously computes a function that represents how "anxious" it is when it is active (switched on) and hence has the potential to get into a hazardous state. An agent based approach allows us to treat each device independently, as well as allowing us to estimate the overall state of the house or any room from the anxieties of each of the devices. In other words, the more devices that are active, each with its own anxiety value, the more anxious the house or room would become. An extreme case would be cooking the dinner and running the bath at the same time. Anxiety would be kept low if the occupant kept moving between the stove and the bath to check the progress of the cooking and bath filling, but become high if the occupant ignored one or both devices. Note the kitchen could have an anxiety based on the anxiety of all the hazardous devices present in the kitchen. If both the stove and fridge are in hazardous states and not attended to, then the kitchen anxiety should be some form of combination of these anxieties.

Devices are defined as monitored objects within the environment, interaction with devices is monitored through the use of methods such as sensors (e.g. reed switches and pressure pads). Devices are considered as either hazardous or passive. A hazardous device refers to devices that have to be attended to while they are in a hazardous state, e.g. stove or fridge. The second class, passive devices, consist of devices for which there is no hazardous state, such as cupboards. For example, no hazardous state is associated with a cupboard as no hazard is introduced by leaving the cupboard open.

A measure of anxiety for each hazardous device is formulated, with the grouping of the anxiety measures for hazardous devices forming *the anxious home*. Anxiety is formulated with statistical models consisting of a model representing typical interactions of a user with the device, and a model representing interaction with other devices, both hazardous and passive, while the device is active. The anxiety for the hazardous device is then determined using a probabilistic framework generated using the expected time periods between an interaction with the hazardous device, and the other devices within the environment. As this approach does not model sequences directly, complex activity patterns along with variations in activity sequences are accounted for. The anxiety is then used as an indicator to determine hazardous situations and provide feedback to a user.

From the above description, it can be inferred that a number of parameters are needed to describe how anxiety works. We take a learning approach to this through interacting with the occupant as initially, only the occupant (or carer) will really understand what is normal and abnormal. In this paper we take a pessimistic approach to anxiety and choose the worst case scenario. Such a scenario can occur when someone is standing in the kitchen for a long period of time or has collapsed on the floor. Given that we are using device activities to infer intent, these two scenarios cannot be separated so we assume the worst — they have collapsed. Pessimistically, if we ask the occupant (or carer) if they are okay (normal) when the anxiety exceeds a threshold, we can use the information about the event to update the parameters.

4. Multi-modal approach to anxiety determination

In experimentation, the anxiety is determined for each hazardous device independently. In the context of modelling the interactions associated with the hazardous device for which the anxiety is being determined, the remaining hazardous devices are considered to be passive.

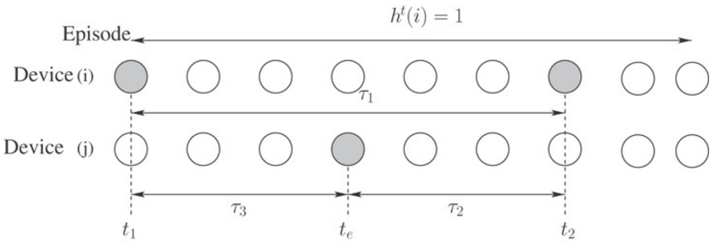


Fig. 1. Anxiety event sequence for device d_i in a hazardous state.

To incorporate audio into the statistical model for anxiety we treat audio activity as an indication of an interaction between the occupant and the environment. That is, the environment is the monitored device, and the audio is the sensor. The presence of the audio is then treated as an interaction with a passive device, indicating a non-hazardous interaction between the user and environment. Therefore the presence of audio activity, i.e. the presence of a segment of foreground audio, is processed in a similar manner to a sensor. The beginning of a foreground segment of audio indicates the activation, or on state, of the audio activity, which then reverts to the *off* state on completion of the activity, the transition from foreground audio to background. We note that the audio is processed as a simple, yet pervasive, sensor. Such a sensor has low processing overheads in conjunction with a number of advantages over visual sensors, such as video and motion sensors. Such advantages arise due to the nature of the propagation of sound signals, which enable the sound sensor to be pervasive, i.e. a single sensor covering a large area. Additionally, the audio sensor does not suffer from occlusion to the extent observed with visual sensors. This use of audio as a pervasive sensor was first proposed in [24].

In this paper P denotes the cumulative distribution function, and p denotes probability density. In determining the anxiety for the hazardous device d_i, we use a number of statistical models. For the device d_i we define the Self Interaction Duration model (SID), $p^{d_i}_{\mathrm{SID}}$, which denotes the probability density distribution of the time intervals between interactions with d_i. The corresponding cumulative distribution is represented by $P^{d_i}_{\mathrm{SID}}$. The closer this probability gets to one, the more anxious device d_i becomes.

To formally derive the probabilities, we consider a hazardous episode for device d_i. The terminology used is as follows; a value of $h^t(i) = 1$ indicates device d_i is in a potentially hazardous state, a value of $d^i(t) = 1$ indicates an interaction with device d_i at time t, and $d^i(t) = 0$ indicates that device d_i was not interacted with at time t. Fig. 1 displays an example time-line of a hazardous sequence for device d_i and the interaction of a user with devices d_i and d_j, where a user interacts with device d_i at times t_1 and t_2 and with device d_j at time t_e, which is within the interval t_1 to t_2.

We define the event E^i_1 as the event that captures the consecutive interactions of duration τ_1 with device d_i, while d_i is in a hazardous state. Event E^i_1 is then defined as,

$$E^i_1(\tau_1) = [d^i(t_1) = 1 \cap d^i(t_2) = 1 \cap \\ ([d^i(t_1+\tau_1) = 0] \cap [t_1 < \tau_1 < t_2])|h^{t_1,t_2}(i) = 1] \quad (1)$$

where $\tau_1 = t_2 - t_1$, and $h^{t_1,t_2}(i) = 1$ indicates that device d_i is in a hazardous state between

t_1 and t_2. That is, E_1^i describes the event device d_i is interacted with at t_1, with the next interaction with d_i occurring at t_2. The $p_{\text{SID}}^{d_i}$ can thus be defined as

$$p_{\text{SID}}^{d_i}(\tau_1) = \frac{\sum_{t_1=0}^{T} E_1^i(\tau_1)}{\sum_{\tau_1} \sum_{t_1=0}^{T} E_1^i(\tau_1)}, \tag{2}$$

where $t_1 = 0$ and T represent the beginning and end of the training sequence respectively.

For each passive device we use, we define two statistical models, the Inter Interaction Duration model (IID), $p_{\text{IID}}^{d_i}$, and the Inter Activity Duration model (IAD), $p_{\text{IAD}}^{d_i}$.

P_{IID} captures the correlation that when passive device d_j is interacted with whilst d_i is in a hazardous state, device d_i will be interacted with again (see Fig. 1). To formulate $p_{\text{IID}}^{d_i,d_j}$, we consider the event $E_2^{i,j}$, which captures the interaction interval of device d_i being interacted with after device d_j, given that d_i is in a hazardous state. Event $E_2^{i,j}$ is defined as:

$$\begin{aligned} E_2^{i,j}(\tau_2) = {} & [d^i(t_2) = 1 \cap d^j(t_e) = 1 \cap \\ & (d^i(t_e + \tau_2) = 0 \cap \tau_2 < t_2) | h^{t_1,t_2}(i) = 1] \end{aligned} \tag{3}$$

where $\tau_2 = t_2 - t_e$ (see Fig. 1), $p_{\text{IID}}^{d_i,d_j}$ is defined as

$$p_{\text{IID}}^{d_i,d_j}(\tau_2) = \frac{\sum_{t_2=0}^{T} E_2^{i,j}(\tau_2)}{\sum_{\tau_2} \sum_{t_2=0}^{T} E_2^{i,j}(\tau_2)}. \tag{4}$$

The corresponding cumulative distribution is represented by the $P_{\text{IID}}^{d_i,d_j}$.

P_{IAD} captures the correlation between a device in the hazardous state and potential interactions with other devices (see Fig. 1). To formulate $p_{\text{IAD}}^{d_i,d_j}$, we consider the event $E_3^{i,j}$, which captures the interaction interval between device d_j being interacted with after device d_i is interacted with, given that d_i is in a hazardous state. Event $E_3^{i,j}$ is defined as:

$$\begin{aligned} E_3^{i,j}(\tau_3) = {} & [d^i(t_1) = 1 \cap d^j(t_e) = 1 \cap \\ & ([d^i(t_1 + \tau_3) = 0] \cap [t_1 < \tau_3 < t_e]) \mid h^{t_1,t_e}(i) = 1] \end{aligned} \tag{5}$$

where $\tau_3 = t_e - t_1$ (see Fig. 1), $p_{\text{IAD}}^{d_i,d_j}$ is defined as

$$p_{\text{IAD}}^{d_i,d_j}(\tau_3) = \frac{\sum_{t_1=0}^{T} E_3^{i,j}(\tau_3)}{\sum_{\tau_3} \sum_{t_1=0}^{T} E_3^{i,j}(\tau_3)}. \tag{6}$$

The corresponding cumulative distribution is represented by the $P_{\text{IAD}}^{d_i,d_j}$.

We also determine the Interaction Event model (IE), $P_{\text{IE}}^{d_i,d_j}$, which denotes the probability of interaction between the user with passive device d_j while device d_i is in a hazardous state. For example, $P_{\text{IE}}^{d_i,d_j} = 0.9$ means that 90% of the time device d_i is in a hazardous state, the user interacts with device j. Consider event $E_4^{i,j}(t)$, defined as:

$$E_4^{i,j}(t) = \left[\left(\sum_t d^j(t)\right) \geq 1 \mid h^t(i) = 1\right]. \tag{7}$$

$P_{\text{IE}}^{d_i,d_j}$ is then defined as

$$P_{\text{IE}}^{d_i,d_j} = \frac{\sum E_4^i(j,t)}{\|h(i)\|}. \tag{8}$$

The anxiety associated with a device d_i is attenuated if a user interacts with devices in the environment associated with device d_i. Consequently, we modify the probability associated with the hazardous device $P_{\text{SID}}^{d_i}$, to reflect these interactions. The scaling factor associated with each device in the environment is defined as

$$\begin{aligned} S^{d_i,d_j}(t, t_1, t_{e_j}) = 1.0 - P_{\text{IE}}^{d_i,d_j} \times (1.0 - P_{\text{IAD}}^{d_i,d_j}(t_{e_j} - t_1)) \\ \times (1.0 - P_{\text{IID}}^{d_i,d_j}(t - t_{e_j})), \end{aligned} \tag{9}$$

where t is the current time, t_1 is the time of last interact with device d_i and t_{e_j} is the time of last interaction with device d_j. The value $1.0 - P_{\text{IID}}^{d_i,d_j}()$ represents the probability that a user will interact with device d_i after the current time t, given that device d_j was interacted with at time t_e. The value $1.0 - P_{\text{IAD}}^{d_i,d_j}()$ represents the probability of interacting with device d_j, at time t_e given an interaction with device d_i, at time t_1.

The probabilities are then incorporated to determine an overall anxiety associated with device d_i:

$$\text{Anxiety}_{\text{overall}}^{d_i}(t) = P_{\text{SID}}^{d_i}(t - t_1) \times \prod_{\forall e_j} S^{d_i,d_j}(\tau_2, \tau_3) \tag{10}$$

where e_j is an event for device d_j and assuming that $e_j \forall j$ are independent of each other.

One problem with this formulation is that if the occupant repeatedly interacts with a device, the anxiety for the hazardous device will keep on reducing. This can be overcome by only using the latest interaction for each device. This can be argued for because once the latest event occurs, all the previous events are not relevant. The anxiety then becomes:

$$\text{Anxiety}_{\text{overall}}^{d_i}(t) = P_{\text{SID}}^{d_i}(t - t_o) \times \prod_{\forall j} S^{d_i,d_j}(t_o, t - t_{e_j}^{\max}) \tag{11}$$

where $t_{e_j}^{\max}$ is the time of the last event for device d_j.

As the anxiety is modelled for each hazardous device independently, the unification of the anxieties need to be considered. Currently the device with the highest anxiety is used to represent the overall anxiety associated with the house. A value for the anxiety of 1.0 indicates that something that has never been seen before has occurred. In keeping with the

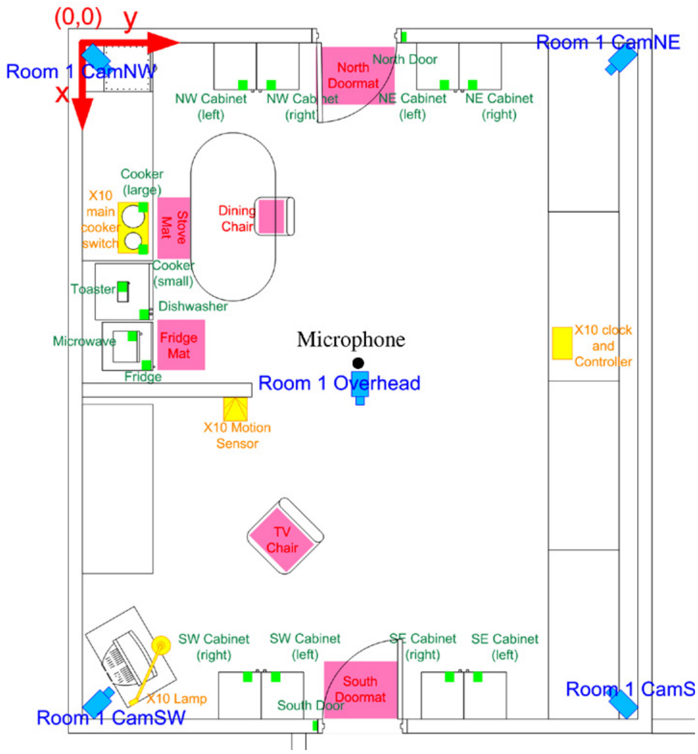


Fig. 2. One room of the Smart House laboratory environment.

pessimistic nature of the anxiety, a lower threshold is set, e.g. 0.8, at which point the user is asked if everything is okay (normal).

5. Experimentation

5.1. Experimental environment

To explore these and other ideas, we developed a Smart House laboratory environment. The laboratory is populated with a number of devices to simulate those that would be found in a typical house. The house has several rooms: a kitchen, lounge and bedroom (Fig. 2 shows two of the rooms). The kitchen includes a small electric stove, microwave oven, fridge, dishwasher, cupboards, a kitchen table and chair. Each device is augmented with sensors to detect interaction by the occupant. Reed switches detect the opening and closing of doors (e.g. the fridge, dishwasher, microwave, cupboards), while pressure mats detect the proximity of the occupant to the doorways. For hazardous devices, pressure mats are positioned on the floor in front of each device to detect proximity and hence potential interaction by the occupant.

An omnidirectional microphone was attached to the ceiling at the centre of the room next to the *Room 1: overhead camera* to capture the audio associated with the activities in the smart house. The audio activity was determined by synchronising the start and end

Table 1
Example of a normal breakfast scenario

	Activity
1	Get ingredients from the fridge
2	Turn on stove and start cooking
3	Make coffee
4	Get cereal and check on stove
5	Eat cereal
6	Put cereal back and check on stove
7	Put ingredient back in the fridge
8	Turn off stove
9	Eat cooked breakfast

time stamps for foreground sections of the audio signal with the logs obtained from the sensor data.

5.2. Activity data

Sensor and audio data were collected for a number of test sequences, consisting of a *normal* scenario in conjunction with a number of abnormal scenarios. To reduce the time required to acquire data, the activities within the scenarios were accelerated.

The normal scenario consisted of 35 sequences depicting activities associated with making breakfast. We focus on the anxiety with respect to the stove. The anxiety increases over time if no interaction with the stove is determined, and reduces to zero upon interaction. Variations were present in both the sequence and duration of events, and the presence of certain events within sequences. The interactions with monitored objects included the fridge, stove, microwave, dishwasher, toaster, and a cabinet. Audio activities present that were not associated with a monitored object were predominantly associated with eating breakfast (e.g. cutlery). Further audio activities occurred due to interaction with the devices monitored by sensors, e.g. doors slamming. Table 1 displays a typical example of a normal scenario.

Seven sequences were generated depicting abnormal scenarios to test the determination of the anxiety for events not seen within the training sequences.

5.3. Audio processing

In total, in excess of four and a half hours of audio data was captured. The normal scenario data was captured in two sessions. Each session consisted of a number of contiguous sequences within a single audio signal. Contiguous capture was necessary to enable the adaptation of the background audio within the environment. The first session consisted of 1.67 h of audio comprising 21 sequences. The second session contained 14 sequences in 1.33 h of audio. The hazardous sequences were captured within a separate contiguous audio sequence of 1.73 h in duration.

The audio signal was captured at 44.1 kHz, 16 bit, mono, wave format, and the background was modelled at a clip size of 1 s. The mean of the wavelet coefficient energy for seven frequency sub-bands was used to characterise the audio signal, and the parameters of the algorithm were adjusted to allow multiple background models. Approximately

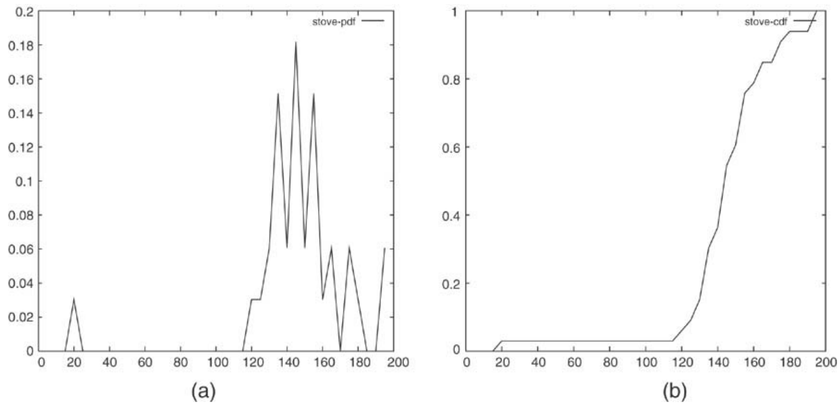


Fig. 3. The probability density function, $p_{\mathrm{SID}}^{\mathrm{STOVE}}$, and cumulative distribution function $P_{\mathrm{SID}}^{\mathrm{STOVE}}$ for the stove (time in seconds).

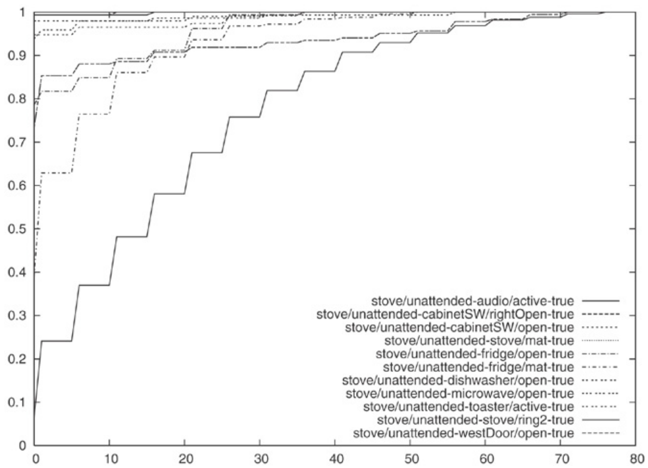


Fig. 4. The cumulative distribution functions for other devices ($P_{\mathrm{IID}}^{\mathrm{STOVE},d_j}$) while the stove is in a hazardous state (time in seconds).

98.4% of the background audio was modelled correctly. This was determined using the number of foreground events detected by the algorithm over a combined total of 25.4 min of inactivity recorded at the beginning and end of each data capture sequence. For the purposes of determining the background accuracy, all audio within the periods of inactivity was considered to be background, including spurious noises occurring outside the smart-house environment.

5.4. Training

To determine the anxiety, 32 normal sequences were used to generate the statistical models, according to Eqs. (1)–(8). Figs. 3 and 4 show example probability density and

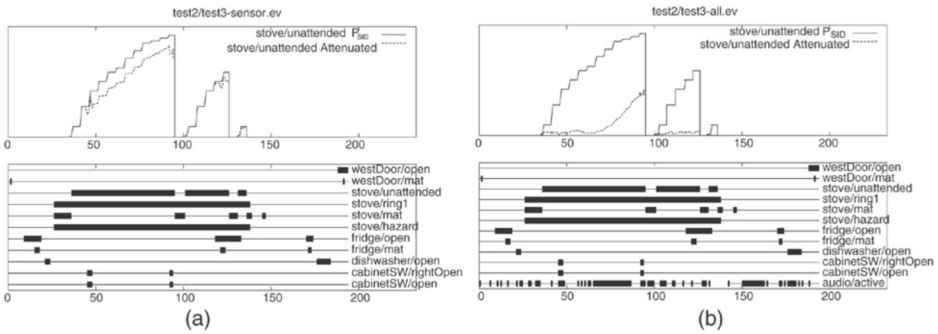


Fig. 5. Test normal sequence 1 (a) without audio (b) with audio.

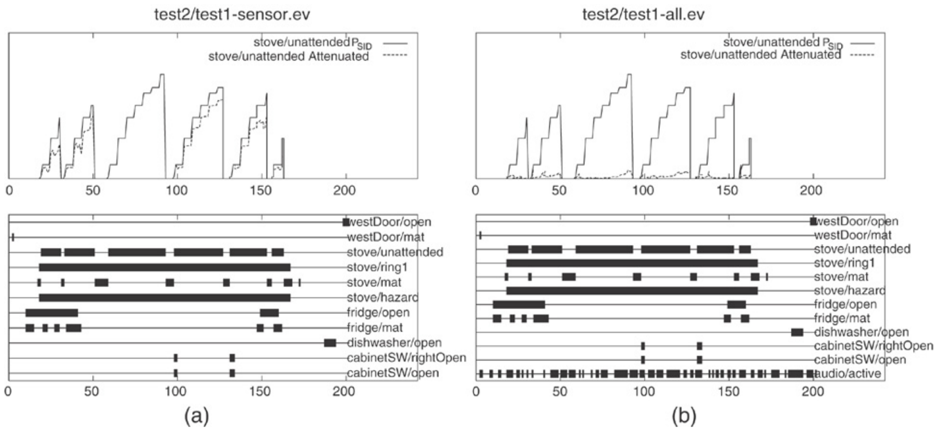


Fig. 6. Test normal sequence 2 (a) without audio (b) with audio.

cumulative distribution functions generated from the normal activity sequences. Fig. 3 shows an example of the p_{SID} and P_{SID} for the stove, while Fig. 4 shows examples of P_{IID} for a number of passive devices that were interacted with while the stove was in a hazardous state. The subsequent models are then used to determine the anxiety for three normal sequences to determine the effect of incorporating audio, and a number of hazardous scenarios. The anxiety for all cases was determined using Eq. (10).

5.5. Results

The anxiety was calculated and updated at a resolution of 1 s. Results for a number of scenarios are shown in Figs. 5–9. The lower graph for each figure displays the time-line for the sequence of events, indicating patterns of interaction for each device. The upper graph displays the anxiety associated with the stove (vertical axis) calculated according to P_{SID} (solid line), and the anxiety attenuated due to interactions with other devices within the environment (dashed line). The horizontal axis corresponds to the timeline (shown in seconds) of the scenario.

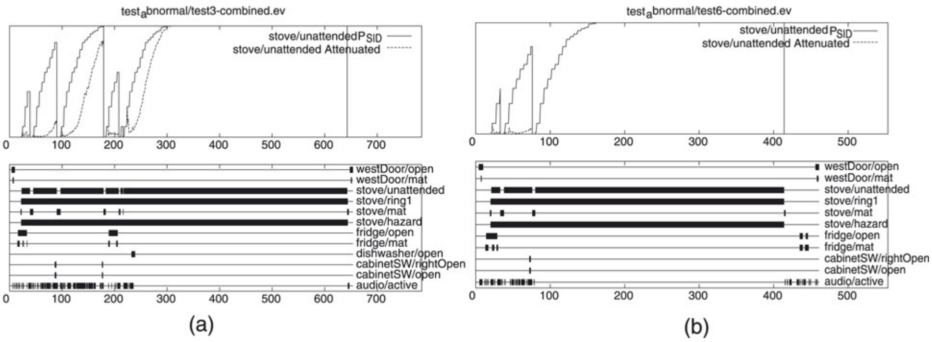


Fig. 7. Abnormal interaction scenario — Absence of activity (a) example sequence 1 (b) example sequence 2.

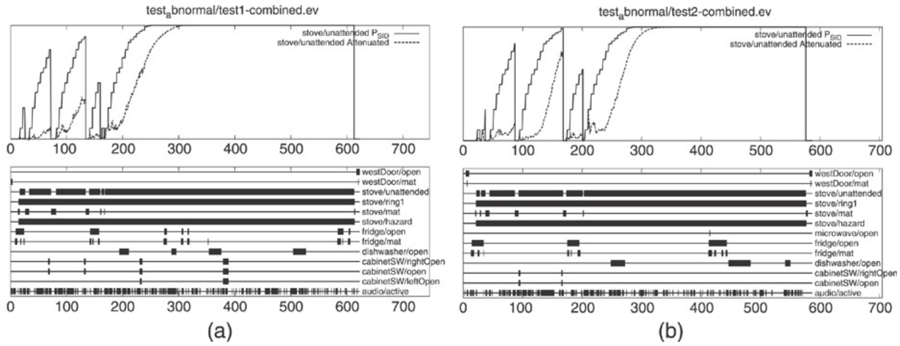


Fig. 8. Abnormal interaction scenario — In presence of activity (a) example sequence 1 (b) example sequence 2.

5.5.1. Attenuating anxiety due to audio

Three normal sequences were used to determine the behaviour of the anxiety for the stove with and without audio to explore how audio enhances the anxiety measure. We test the attenuation of anxiety due to the presence of audio activity.

Figs. 5 and 6 show two normal test sequences, calculating the anxiety both with and without the audio activity data. In Fig. 5(a), the anxiety of the stove when no other device interactions are considered (P_{SID}) and the anxiety when device interactions are taken into account (dashed line) are shown. The user switches the stove on at 36 s, and anxiety starts to increase until 95 s, at which time the user interacts with the stove, reducing the anxiety to 0. The user keeps interacting with the stove until 101 s, as a result of which the anxiety stays at 0. The anxiety then rises from 101–126 s when the user moves away to do other tasks before returning to the stove (126–131 s), again reducing the anxiety to 0. The anxiety rises again until 136 s, at which point the user turns the stove off. The dashed line displays the attenuated anxiety. At 45 s and 92 s the anxiety is attenuated due to interactions with the cabinet. Fig. 5(b), shows the same sequence but this time the audio is included. Whilst the unattenuated anxiety (solid line) behaves as described before, the attenuated anxiety (dashed line) shows that the anxiety has been greatly attenuated as compared to Fig. 5(a), because of the contextual audio.

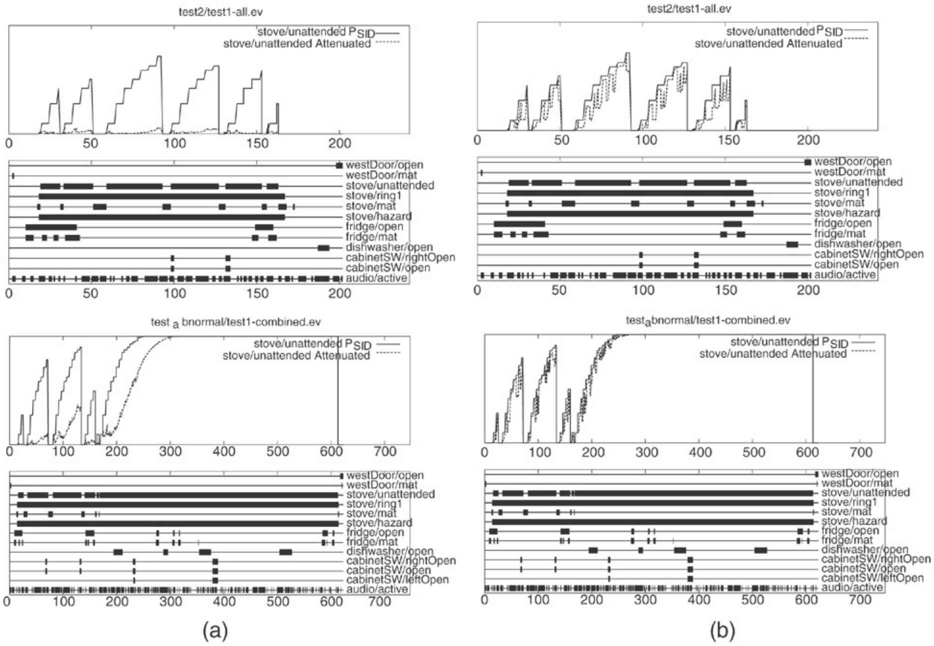


Fig. 9. Normal and abnormal activity sequences for (a) Eq. (10) (b) Eq. (11).

Fig. 6(a) and (b) show the results for a different sequence. In Fig. 6(a), between 59 and 92 s, the lack of interaction with monitored devices causes the anxiety to rise unattenuated. Fig. 6(b) displays a greater level of attenuation due to the presence of audio activity within this section of the sequence.

Both figures display variations of the breakfast scenario, evidenced by the differing interactions with the stove. Despite the differences in the sequence of the activities, the anxiety still produces meaningful results.

From both figures it is evident that the audio activity results in a higher attenuation of the anxiety in comparison with just using the sensor data, as would be expected as the person is in the room. The degree of audio activity present is indicated by the time-line (lower graph). For both figures, the audio occurs more frequently in comparison with the remaining monitored devices.

5.5.2. Abnormality: Absence of activity

The anxiety was determined for four abnormal scenarios characterised by a lack of activity while the stove was in a hazardous state. Fig. 7 shows the anxiety, determined using the combined audio and sensor data, for two examples of the abnormal scenarios where activity is absent.

Fig. 7(a) depicts a normal breakfast scenario from 0 to 225 s with the last interaction with the stove at 216 s. The user subsequently leaves the room without turning off the stove. Note the absence of audio as the user left the room. The P_{SID} of the stove rises to

1.0 at 299 s, with the attenuated anxiety reaching 1.0 at 311 s due to the user interacting with the environment after the last interaction with the stove, from 216 to 225 s.

Fig. 7(b) depicts a break in the normal breakfast scenario, at 80 s, characterised by an absence of activity without completing the breakfast sequence, e.g. the user collapsed and no longer made any noise. As no activity is present from the last interaction with the stove, the anxiety raises to 1.0 at 162 s, without any attenuation effect.

In both cases the lack of audio activity meant the anxiety rose to 1.0, meaning an alarm would be raised.

5.5.3. Abnormality: Presence of activity

The anxiety was then determined for three abnormal scenarios characterised by the presence of activity within the room, and a lack of interaction with the stove while it was in a hazardous state. Fig. 8 shows the anxiety, determined using the combined audio and sensor data, for two examples of the abnormal scenarios where activity is present.

Fig. 8(a) depicts the normal breakfast scenario from 0 to 168 s, with 168 s being the time of last interaction with the stove. The user subsequently forgets to turn off the stove after cooking, but remains active within the room, both interacting with sensors and producing audio activity. The P_{SID} reaches 1.0 at 249 s, and the attenuated anxiety reaches 1.0 at 302 s. Fig. 8(a) displays a similar scenario, the P_{SID} reaches 1.0 at 285 s, and the attenuated anxiety reaches 1.0 at 380 s.

In this case, it is the absence of interaction with the stove that results in an increase in the activity due to the use of the P_{IAD} distribution. The presence of activity within the environment results in an increased time for the anxiety to reach a value of 1.0.

5.5.4. Anxiety attenuation through events or devices

Eq. (10) accounts for all interactions of a user with the environment, while Eq. (11) only accounts for the last interaction with each device. Fig. 9 shows the results for the calculation of the anxiety using both methods for a normal scenario and the abnormal scenario used for Fig. 8(a). The statistical model used for Eq. (11) was generated using the last event for each device immediately prior to the interaction with the hazardous device.

From the figure it can be seen that the inclusion of only the recent history for each device results in a reduced attenuation of the anxiety in comparison with taking the recent history of the activity of the user. There is less attenuation of the anxiety, along with a reduction of the time period over which the attenuation occurs. For example, for the abnormal sequence the attenuated anxiety reaches a value of 1.0 at 270 s, in comparison with 312 s for the method associated with Eq. (10). This is due to the reduction in the expected time between interacting with a passive device and subsequently interacting with the hazardous device that is inherent in the generation of the statistical model.

Consequently the method associated with Eq. (11) is more appropriate when a more sensitive indication of anxiety is required. Eq. (10) results in a lower overall anxiety for sections of the sequences that are within the bounds of normality given the observed normal sequences.

5.6. Discussion

In this section we examine the results obtained from applying the anxiety model to both normal and abnormal scenarios. For normal activity associated with the stove, accounting for interaction with the environment successfully attenuated the anxiety. This was especially prominent when accounting for audio, which resulted in an increased attenuation of the anxiety as more environmental interactions were accounted for. The attenuation of the anxiety reduces the number of false alarms raised by the system, indicated by the occupant being active in the environment, and subsequently interacting with the stove.

Two types of abnormal interaction with the stove were examined. The first was the absence of activity either immediately or shortly after interacting with the stove, e.g. the occupant experiences a fall. In this case the absence of activity after interacting with the stove results in either no attenuation of the anxiety, or a short delay in reaching an anxiety value of 1.0 (12 s). The second abnormal activity corresponded to the presence of activity in the absence of an interaction with the stove, e.g. the occupant forgets to turn off the stove. In this case, the presence of activity in the environment by the user results in the attenuation of the anxiety, with a delay in reaching a value of 1.0. In both abnormal interaction cases the alarm would be correctly raised. The delay in raising the alarm in the second case is acceptable as the occupant is still active in the environment.

6. Feedback implementation

The link between the anxiety system and the user is completed through the use of a personal digital assistant (PDA). Such a method is necessary due to the pessimistic nature of the anxiety, requiring user interaction once the anxiety of the environment climbs above a threshold.

The system uses a server for most of the processing with a back-end database to store parameters, state of the house etc. and is accessible via the web for configuration, querying by a carer and other activities.

A PDA is used for cognitive support which is wirelessly connected to the server so the occupant is free to roam around the house. The server alerts the occupant via the PDA about any hazards and expects the occupant to respond promptly. Lack of a prompt reply alerts the server to a problem and the carer is alerted.

A server takes input from the sensors in the smart house and interacts with the web and the PDA. All software for the server and PDA is written in Java for portability. The PDA talks to the server via WiFi with the network layer on the server and PDA with messages passed between the PDA and the server via object serialisation. A priority queue is used so that important messages from the PDA can be processed immediately. The network layer ensures continued connection between the server and client by keeping in constant contact. If this contact is broken, it implies that the connection has been lost and the occupant is out of range or something untoward has happened. The web-based interface uses a SQL database as the back end. Features of the interface include the display of the real-time state of the sensors in the house, log of events, parameters for the sensors and the means for creating and modifying reminders for the occupant and carer. As the web-based interface is only for authorised users, it includes user account management facilities. The PDA incorporates the popular alarm facility in which a button on the PDA is used for calling for help. No interaction of the occupant with the device will eventually lead to the occupant being reminded, via the PDA, that they have left the device unattended, and eventually, if no action is taken, the device is turned off by the house and the carer notified.

The PDA extends the anxiety model as it closes the loop on the system querying the occupant when it becomes anxious. Once the anxiety of the house reaches a certain level the system causes the PDA to signal the occupant audibly and visually and displays a message on the screen describing the hazard and whether this is okay for the occupant or not. If the occupant responds “yes” the anxiety of the system is lowered for that device. If the occupant responds “no”, or there is no response at all, the device is turned off and the carer notified. The PDA is also given an anxiety value. If the occupant doesn’t respond to an alert, the anxiety of the PDA rises eventually making the system alert the carer.

7. Conclusion

In this paper we have proposed a method for hazard detection in smart environments using a fusion of multi-modal data within an emotive computing framework.

Previous approaches to determining abnormal activity have centred around activity recognition fusing data collected from various sensors. We approach the problem from a different perspective. Rather than using activity recognition, we determine normality with respect to the patterns of interaction associated with each hazardous device. We use a probabilistic approach that enables the modelling of complex interactions without being reliant on the sequence of interactions.

The significance of this work lies in two key areas. Firstly, a scalable, agent based method for detecting hazards in a smart house environment is proposed. Secondly, we integrate multi-modal data from a number of sources into the anxiety framework. In particular, the inclusion of audio as a pervasive sensor extends audio analysis to the field of surveillance and monitoring. Audio is a powerful cue that can be mapped to higher level semantic analysis.

The advantage of audio analysis in comparison with simple sensors lies in the contextual and pervasive nature of the audio data. Audio also offers an advantage over video analysis due to the lower processing overheads. Audio background modelling classifies background audio according to the dominant characteristics of the audio over a period of time. Foreground audio classification is therefore determined by a difference in the audio signal from the background. The background is adaptively modelled on-line, enabling the determination of the foreground sounds across varying and changing background audio. We argue that the novel sounds, i.e. the foreground, are sounds associated with an activity. This links the audio with a higher level semantic meaning.

The anxiety was determined for a number of normal and hazardous sequences, producing meaningful results in both cases. The inclusion of the audio activity resulted in a more meaningful attenuation of the anxiety as sounds are made by people even though they are not interacting with a monitored device.

The anxiety represents a pessimistic emotive model, which is used to determine when an occupant should be reminded of an ongoing hazard within the environment. The response can be used to manage the hazard, and to refine the anxiety model. Due to the probabilistic and pessimistic approach, the parameters are learned incrementally for a particular person and house by querying the occupant each time the anxiety threshold for a hazardous agent is exceeded. As such, the inclusion of the PDA is crucial in training the anxiety model. Initially the number of false alarms generated would be high. As each false alarm prompts the user for a response via the PDA, normal interactions with hazardous agents are learned, and over time the number of false alarms would fall.

Future work includes determining abnormal interaction with the environment given spatial and temporal context. For example, the concept of anxiety could be used to determine if the occupant has been in the bathroom for an unusually long period of time, without interacting with the environment, given the time of day. When employed in such a manner, the anxiety can be extended to determine unusual events, or interaction with the environment, such as the occupant experiencing a fall.

References

[1] A.P. Glascock, D.M. Kutzik, Moving telematics from the laboratory to a truly enabling technology within the community, in: Proc. ICOST'2004: 2nd. Int. Conf. on Smart Homes and Health Telematics, Singapore, 2004, pp. 145–153.

[2] H.H. Bui, S. Venkatesh, G. West, Policy recognition in the abstract hidden Markov model, Journal of Artificial Intelligence Research 17 (2002) 451–499.

[3] N.T. Nguyen, S. Venkatesh, G. West, H.H. Bui, Hierarchical monitoring of people's behaviours in complex environments using multiple cameras, International Conference Pattern Recognition.

[4] X. Bian, G. Abowd, J. Rehg, Using sound source localization in a home environment, in: Proceedings of The 3rd International Conference on Pervasive Computing, Munich, Germany, May, 2005, pp. 19–25.

[5] G. West, S. Greenhill, S. Venkatesh, A probabilistic approach to the anxious home, in: The 29th Annual International Computer Software and Applications Conference — COMPSAC 2005, Edinburgh, Scotland, July, 2005.

[6] R.W. Picard, Affective Computing, MIT Press, 2002.

[7] M. Azlan, I. Cartwright, N. Jones, T. Quirk, G. West, Multi-modal monitoring of the aged in their own homes, in: Proc. ICOST'2005: 3rd. Int. Conf. on Smart Homes and Health Telematics, Magog, Canada, July, 2005.

[8] E.M. Tapia, S.S. Intille, K. Larson, Activity recognition in the home using simple and ubiquitous sensors, in: Proc. PERVASIVE 2004, in: Lecture Notes in Computer Science, vol. 3001, Springer-Verlag, 2004, pp. 158–175.

[9] M. Chan, E. Campo, D. Esteve, Monitoring elderly people using a multisensor system, in: Proc. ICOST'2004: 2nd. Int. Conf. on Smart Homes and Health Telematics, Singapore, 2004, pp. 162–169.

[10] K. Matsuoka, Aware home understading life activities, in: Proc. ICOST'2004: 2nd. Int. Conf. on Smart Homes and Health Telematics, Singapore, 2004, pp. 186–193.

[11] A.P. Glascock, D.M. Kutzik, The impact of behavioural monitoring technology on the provision of health care in the home, Journal of Universal Computer Science 11 (1) (2006) 59–79.

[12] A.D. Joseph, Successful aging, IEEE Pervasive Computing 3 (2) (2004) 48–50.

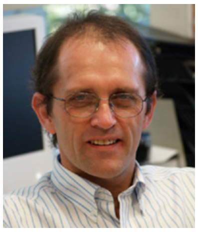
[13] H.H. Bui, S. Venkatesh, G. West, Tracking and surveillance in wide-area spatial environments using the abstract hidden Markov model, International Journal of Pattern Recognition and Artificial Intelligence 15 (1) (2001) 177–195.

[14] S. Lühr, H.H. Bui, S. Venkatesh, G.A.W. West, Recognition of human activity through hierarchical stochastic learning, in: IEEE International Conference on Pervasive Computing and Communications, Texas, USA, 2003.

[15] N.T. Nguyen, H.H. Bui, S. Venkatesh, G. West, Recognising and monitoring high-level behaviours in complex spatial environments, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Wisconsin, USA, 2003.

[16] M. Vacher, D. Istrate, L. Besacier, J.F. Serignat, E. Castelli, Life sounds extraction and classification in noisy environment, in: 5th IASTED-SIP, ACTA Press, Calgary, Hawaii, 2003.
[17] M. Vacher, D. Istrate, L. Besacier, J.F. Serignat, E. Castelli, Sound detection and classification for medical telesurvey, in: 2nd Conference on Biomedical Engineering, Innsbruck, Austria, ACTA Press, Calgary, 2004, pp. 395–398.
[18] J. Chen, A.H. Kam, J. Zhang, N. Liu, L. Shue, Bathroom activity monitoring based on sound, in: Pervasive Computing, Munich, Germany, 2005, pp. 47–61.
[19] D.P.W. Ellis, Detecting alarm sounds, in: Consistent and Reliable Acoustic Cues for sound analysis, Aalborg, Denmark, 2001.
[20] M. Cowling, R. Sitte, Comparison of techniques for environmental sound recognition, Pattern Recognition Letters 24 (15) (2003) 2895–2907.
[21] M. Cowling, Non-speech environmental sound classification system for autonomous surveillance, Ph.D., Griffith University, 2004.
[22] A. Härmä, M. McKinney, J. Skowronek, Automatic surveillance of the acoustic activity in our living environment, in: IEEE International Conference on Multimedia and Expo, ICME 2005, Amsterdam, Netherlands, 2005.
[23] S. Moncrieff, S. Venkatesh, G. West, Persistent audio modelling for background determination, in: IEEE International Conference on Multimedia and Expo, ICME 2005, Amsterdam, Netherlands, 2005.
[24] S. Moncrieff, S. Venkatesh, G. West, S. Greenhill, Incorporating contextual audio for an actively anxious smart house, in: Intelligent Sensors, Sensor Networks and Information Processing Conference (Proceedings of the 2005 International Conference, Melbourne, Australia, 2005). 2005, pp. 373–378.

Svetha Venkatesh is a Professor at Curtin University of Technology's Department of Computing. Her research interests are large-scale pattern recognition, image understanding and applications of computer vision to image and video indexing and retrieval. She co-directs the university's Centre of Excellence in Intelligent Operations Management and directs the university's Institute of Multi-Sensor Processing and Content Analysis.



Geoff West is a Professor of Computer Science at Curtin University of Technology where he serves as the head of the Department of Computing. His research interests include 3D object recognition, feature extraction, surveillance, smart homes, pervasive computing, small-scale systems, automatic visual inspection, data mining, telemedicine and related applications. He received a Ph.D. in Systems Engineering from City University, London. He is a senior member of the IEEE, a member of the IEE (UK), a fellow of the IEAust, and a Chartered Engineer.