# Examining the Utility of Affective Response in Search of Personal Lifelogs

Liadh Kelly
Centre for Digital Video Processing
Dublin City University
Ireland
lkelly@computing.dcu.ie

Gareth J.F. Jones
Centre for Digital Video Processing
Dublin City University
Ireland
gjones@computing.dcu.ie

## ABSTRACT

Personal lifelog archives contain digital records captured from an individual's daily life, for example emails, documents edited, webpages downloaded and photographs taken. While capturing this information is becoming increasingly easy, subsequently locating interesting items from within these archives is a significant challenge. One potential source of information to identify items of importance to an individual is their affective state during the capture of the information. The strength of an individual's affective response to their current situation can often be gauged from their physiological response. For this study we explored the utility of the following biometric features to indicate significant items: galvanic skin response (GSR), heart rate (HR) and skin temperature (ST). Significant or important events tend to raise an individual's arousal level, causing a measurable biometric response. We examined the utility of using biometric response to identify significant items and for re-ranking traditional information retrieval (IR) result sets. Results obtained indicate that skin temperature is most useful for extracting interesting items from personal archives containing passively captured images, computer activity and SMS messages.

## 1. INTRODUCTION

Advances in digital technologies mean that a wealth of personal information is now becoming available in digital format. This information can be gathered together and stored in a personal lifelog [2] [7] [12]. Personal lifelog archives can contain everything from items read, written, or downloaded; to footage from life experiences, e.g. photographs taken, videos seen, music heard, details of places visited, details of people met, etc, along with details of location and social context. Finding important relevant items from within these archives in response to user queries, or presenting interesting data to a subject browsing through their archive, poses significant challenges. Any additional information which can assist in identifying important items is thus potentially very important. Such information could be used in the re-ranking of information retrieval (IR) result sets, and for the promotion of interesting items when browsing a lifelog collection. One potential source of useful information is the user's biometric response associated with an item. In this study we explore three biometric responses associated with items, namely galvanic skin response (GSR), heart rate (HR) and skin temperature (ST).

Previous work has shown an individual's biometric response to be related to their overall arousal levels [13]. Significant or important events tend to raise an individual's arousal level, causing a measurable biometric response [16]. Events that can be recalled clearly in the future are often those which were important or emotional in our lives [6]. It has been demonstrated that the strength of the declarative or explicit memory for such emotionally charged events has a biological basis within the brain. Specifically involving interaction between the amygdala and the hippocampal memory system [5]. Variations in arousal level elicit physiological responses such as changes in heart rate or increased sweat production. Thus one way of observing an arousal response is by measuring the skin conductance response (SCR) (also referred to as the galvanic skin response (GSR)). The GSR reflects a change in the electrical conductivity of the skin as a result of variation in the activity of the sweat glands. It can be measured even if this change is only subtle and transient, and the individual concerned is not obviously sweating [6]. Curiously these biometric responses can be anticipatory of the consequences of a possible potentially risky action based on previous experiences, and may be observed before an individual is even consciously aware that the action may have significant consequences [4]. Arousal response can also be observed through skin temperature. With increased arousal levels, sympathetic nervous activity increases, resulting in a decrease of blood flow in peripheral vessels. This blood flow decrease causes a decrease in skin temperature [21]. Current technologies enable the capture of a number of biometric measures on a continuous basis. For example using a device such as the BodyMedia SenseWear Pro II armband [3] which can continuously record the wearer's GSR and skin temperature, or using the Polar Heart rate Monitor [19] which can continuously record the wearer's heart rate.

We propose that items or events which are important to an individual at the time they occurred may be useful to the individual again in the future, and further that such incidents are associated with emotional responses that can be detected by measuring an individual's biometric response when experiencing these events. Thus recording GSR, HR


*5th Workshop on Emotion in HCI, British HCI Conference 2009, September 1, 2009,* Cambridge, UK.

and ST as part of a lifelog may enable us to identify important items or events in a lifelog which would be most interesting for an individual to browse through in the future or which would be most important in a given information searching task.

In this paper we explore our hypothesis and report our findings to date which may guide future research in this area. We describe two studies: the first a study designed to explore the use of biometric data in detecting useful lifelog items for future browsing; the second is a study to investigate the utility of biometric response in re-ranking traditional information retrieval result lists. The next section describes the test-set gathered for these experiments. Section 3 presents our first study setup and results obtained. Our second study and results obtained are provided in Section 4. We then conclude the paper with a discussion of the findings and directions for future work.

## 2. TEST-SET

In order to explore our hypothesis, a suitable test-set must be available. As part of our ongoing work on personal lifelogs we are gathering long term lifelog collections from a small group of subjects. For the current investigation we augmented these for 3 subjects, for a 1 month period, with capture of their GSR, HR and ST data. For our current experiment we chose to examine: 1) whether these biometric data types can be useful in identifying important and memorable items which the subject may wish to view again in the future; and 2) the utility of these biometric data types in re-ranking the output of a user query driven IR result list. The lifelogs used for these experiments contained computer items accessed (web pages viewed, files created or accessed, emails sent and received, etc), SMSs sent and received, and images capturing an individual's activity.

The heart rate data was collected using a Polar Heart Rate Monitor [19]. The heart rate monitor is worn around the chest, and heart rate readings are transmitted to a watch worn on the subject's wrist. Software provided with the device generates reports, graphs and text files of the heart rate readings for data analysis. All other biometric data was collected using a BodyMedia SenseWear Pro2 armband [3]. The BodyMedia armband is worn on the upper arm and measures a range of psychological data. Data captured includes galvanic skin response along with transverse acceleration, longitudinal acceleration, heat flux, skin temperature, and energy expenditure. Software provided with the device generates graphs, reports and comma separated files of all the sensor output for data analysis. Our three subjects wore the heart rate monitor and BodyMedia devices for a one month period to capture biometric data for this investigation.

In addition to the biometric data, our experimental lifelogs contained data of computer activity, SMSs sent and received, and a visual log of activities. Computer activity was recorded using the Slife package [24]. Slife monitors computer activity and records the event of a window being brought to the foreground. For each event it records: type of application (e.g. web, chat), document source (e.g. Microsoft Word), window title and begin and end time of the event. Window title, application and document source were used to determine extension type (e.g. pdf, doc). The textual content inside the window (e.g. the text of an email, web page or document being written) and path to each file were obtained using MyLifeBits [17] for two subjects who are users of Windows XP and using locally written scripts for the other one who uses Mac OS X. The subjects all used Nokia N95 mobile phones [18] to capture SMSs. Logs of SMSs sent and received were generated using scripts installed on N95s. A visual log of subjects' activities was created using a Microsoft Research SenseCam [8] [23]. The SenseCam is a digital camera, with fish-eye lens, worn around a subject's neck. This passively captures images approximately every 20 seconds. Image capture is triggered based on changes in sensor data captured by the device. For example, high acceleration values, passive infrared (body heat detector) as someone walks in front of the wearer or changes in light level. If no sensor has triggered an image to be captured, the camera takes one anyway after a period of approximately 30 seconds. When worn continuously, roughly 3,000 images are captured in an average day.

Lifelog items (i.e. computer items accessed, SMS sent and received, and SenseCam images) were also annotated with the following types of context data:

- Using time and date information functions were written to determine, the month, day of week, part of week e.g. weekend or weekday, hour, minute, second, and period of the day e.g. morning, afternoon, evening, night, in which the event took place.
- Events were annotated with geo-location. GPS data, wireless network presence and GSM location data was captured by constantly running the Campaignr software, provided to us by UCLA (USA) [9], on subjects N95 mobile phones. From which geo-location was derived using in-house scripts.
- Events were annotated with light status and weather conditions derived using date, time and geo-location information.
- People present were annotated to events. The Campaignr software also recorded co-present Bluetooth devices, from which people present can be uncovered [11], [14].
- Using mobile phone call logs, generated using freeware [15] installed on N95s, events were annotated with phone conversations which occurred when the event took place.

The context data types used in this study were: file name; extension type; date; time; month; day of week; weekday or weekend; morning, afternoon, evening or night; light status.

Lucene [1], an open source search engine, was used to index items and their associated context data into different fields (e.g. day of week field, etc). The StandardAnalyzer built into Lucene was used to index the content of items. This tokenizes the content based on a sophisticated grammar that recognises email address, acronyms, alphanumeric and more; converts to lowercase, and removes stopwords.

## 3. EXPERIMENT 1: LOCATING INTERESTING EVENTS

In this section we describe the setup of a study to examine whether biometric data can be useful in identifying important and memorable events (an event is a group of SenseCam images or computer and SMS items) which the subject may wish to view again in the future. The results of this study suggest that biometric data can be useful in detecting important lifelog events and highlights the types of events it is most beneficial for. We begin by describing the test-set used for this study and the experimental approach taken. The results obtained are then discussed.

## 3.1 Extracting Important Events

We postulate that important events from a lifelog archive are coincident with maximum observed GSR and HR readings and with minimum observed ST readings at the original time of event occurrence, and that these events would be most interesting for subjects during future archive browsing. In this study we investigate three types of biometric reading: GSR, HR and ST. The BodyMedia device samples the values from its inbuilt sensors at settable predefined intervals. Based on results from initial calibration experiments we set the device to capture GSR data once per second and ST data once every ten seconds. The maximum rate of HR data captured afforded by the Polar heart rate monitor device was once every five seconds.

Variations in biometric response occur all the time and can be caused by many things, for example changes in arousal level or changes in physical activity such as walking down a corridor or running. A problem in analysis of biometric data for the purposes of this experiment is to identify variation in biometric data which are likely to be the result of variations in arousal levels, as opposed to physical reasons. An additional source of data that can be inferred from captured biometric data using the BodyMedia armband is the energy expenditure (sample rate set to once per minute) of the individual. Energy expenditure correlates well with periods of physical activity. Thus measured energy expenditure can be used to differentiate between high GSR and HR biometric data levels and low ST biometric data levels, resulting from physical activity and those arising from events experienced from the environment. GSR, HR and ST data captured during periods of energy expenditure above the average energy level * $\alpha$ ($\alpha$ = empirically determined scalar constant) were removed from the data set. To determine correlation between item importance and GSR, HR or ST, we attempted to extract 10 max, 10 average and 10 min [1] GSR lifelog items/events; 10 max, 10 average and 10 min HR items/events; and 10 max, 10 average and 10 min ST lifelog items/events, this corresponds to 5 SenseCam and 5 computer/SMS items/events for each GSR, HR and ST level from each subject's lifelog. The procedure for extraction of these SenseCam and computer/SMS items/events was as follows:

1. Determining begin and end timestamps of max GSR and HR: Begin and end timestamps for periods in a subject's GSR/HR dataset where the GSR/HR level was greater than a preset threshold for an empirically determined number of seconds were recorded. (threshold = average of GSR/HR data * $\beta$, $\beta$ = empirically determined scalar constant)

   Determine begin and end timestamps of max ST: Timestamps were obtained by taking periods where ST levels were less than a preset threshold for an empirically determined number of seconds. (threshold = average of ST data / $\beta$, $\beta$ = empirically determined scalar constant)

   Determining begin and end timestamps of min GSR and HR: Timestamps were obtained by taking periods where GSR/HR levels were less than a preset threshold for an empirically determined number of seconds. (threshold = average of GSR/HR data / $\chi$, $\chi$ = empirically determined scalar constant)

   Determining begin and end timestamps of min ST: Begin and end timestamps for periods in a subject's ST dataset where the ST level was greater than a preset threshold for an empirically determined number of seconds were recorded. (threshold = average of ST data * $\chi$, $\chi$ = empirically determined scalar constant)

   Determining begin and end timestamps of average GSR, HR or ST: Timestamps were obtained by taking periods where GSR/HR/ST levels were greater than threshold1 and less than threshold2 for an empirically determined number of seconds. (threshold1 = average of GSR/HR/ST data - $\delta$, $\delta$= empirically determined scalar constant; threshold2 = average of energy expenditure data + $\sigma$, where $\sigma = \delta$)

2. Extracting items/events from the subject's lifelog: The begin and end timestamps from step 1 were used to extract SenseCam, and computer/SMS events as follows: *if* computer or mobile activity occurred between the begin and end timestamps, these items were extracted, *else if* SenseCam images occurred between the begin and end timestamps, these images were extracted.

3. Removing duplicates: An item/event may cause a max biometric response on one access to it and on a different access cause an average or min biometric response for example. Items were removed from all but their highest occuring threshold (e.g. if a computer file currently in the min collection to be presented to subject caused a max biometric response on a different occasion it was removed from the min collection).

[1] Max GSR/HR = periods of high GSR/HR; max ST = periods of low ST; average GSR/HR/ST = periods of average GSR/HR/ST; min GSR/HR = periods of low GSR/HR; and min ST = periods of high ST.

On completion of this process, having expanded thresholds as far as possible, we had sets of <=45 SenseCam events and <=45 computer/SMS events from each subject's lifelog. These sets of events were used to test our hypothesis, as described in the next section.

## 3.2 Experiment Procedure

The goal of this research is to establish if max periods of GSR, HR and ST are good indicators of lifelog items/events which are most useful for presentation to subjects when browsing their personal information archives. Personal lifelog items of varying GSR, HR and ST were presented to subjects and questionnaires completed to determine if GSR, HR and ST corresponded with memorable-ness, significance of events and desire to view again. Post questionnaire interviews were then conducted. This section describes the details of these procedures.

We wished to establish the relationship between biometric response at time of item/event creation/access on subjects' desire to re-view lifelog items/events over the long-term. We thus waited for nine months after the test-set collection period to present subjects with a set of events taken from their lifelogs. A total of <=90 events generated using the technique described in Section 3.1 were presented to subjects in this set. The set included: for each of GSR, HR and ST <=5 computer or mobile phone activity events with the max GSR/HR/ST and <=5 SenseCam image events corresponding to the max GSR/HR/ST; and for comparison purposes similar sets of events with average GSR, HR, ST and min GSR, HR, ST. For each of average GSR, HR and ST the <=5 computer or mobile phone activity events and <=5 SenseCam events closest to the subjects average GSR, HR and ST were chosen (as described in previous section). For each of min GSR, HR and ST the <=5 computer or mobile phone activity events and <=5 SenseCam events closest to the subject's lowest min GSR/HR/ST were chosen (also described in previous section).

Each subject was presented with their set of <=45 computer/SMS events and <=45 SenseCam events. Subjects were aware that the sets presented to them contained events with varying associated biometric levels and of the specific hypothesis we wished to test. However, they were not aware of the biometric response associated with each event. The questionnaire was explained to subjects and sample answers provided.

The subjects completed Questionnaire 1 for these <=90 events (and returned the completed questionnaire to the investigator). Details of this questionnaire were as follows:

1. Is this event memorable? (4-point scale: 4 = very memorable, 3 = memorable, 2 = not very memorable, 1 = not memorable).

2. Was the event important to you at the time? (4-point scale: 4 = very important, 3 = important, 2 = not very important, 1 = not important).

3. Is the event important to you now? (4-point scale: 4 = very important, 3 = important, 2 = not very important, 1 = not important).

4. Is it interesting to see this event again? (4-point scale: 4 = very interesting, 3 = interesting, 2 = not very interesting, 1 = not interesting).

5. Would you want to view this event again? (3-point scale: 1 = yes, 2 = maybe, 3 = no).

The following sections discuss the findings of this study.

### 3.3 Experiment Results

For analysis purposes, for each of questions 1-4 we took a binary split of the 4-point scale; that is a score of 4 or 3 was taken as positive and a score of 2 or 1 was taken as negative. We then calculated the average number of positive scores for each question for SenseCam and for computer/mobile phone activity events. Figures 1 and 2 show the average number of positive scores for questions 1-4 across the 3 subjects.

These results suggest a certain level of correlation between memorableness, importance at time and GSR and ST levels associated with SenseCam images (Figure 1). However as can be seen in the graph, while memorableness and importance at time of events were well captured by max GSR, these factors did not correlate with perceived current importance or desire to view the images now. During informal interview many reasons for this were uncovered, for example an event, such as an important work meeting while well remembered and important at the time to a subject no longer held any relevance and the subject had no interest in viewing the images relating to this event. Good correlation was observed between ST levels associated with images for subjects perception of current importance of the images and for their interest in viewing them now. Suggesting that ST at time of passive image capture might be a better indicator than GSR of images to present a subject with in the future. The results observed for HR levels were very poor (Figure 1). These results might be partly explained by the unreliability of the HR recording device, which sporadically gave unreliable readings. While efforts were made during experimentation to remove HR readings that appeared erroneous, it is possible that some incorrect readings remained. Other devices can capture HR more reliably, we hope to explore their use in future experiments.

Question 5 on the questionnaire specifically examined if subjects anticipated wanting to view these images again in the future. Results are presented in Figure 3. Overall subjects would or might want to retrieve (yes and not sure in Figure 3) 40% of max GSR response images, this compares with 29% of average GSR response images and 20% of min GSR images. ST levels was a more reliable indicator of subjects possible desire to view images in the future. They would or might want to retrieve (yes and not sure in Figure 3) 45% of max ST images, this compares with 15% of average ST images and 13% of min ST images. Again poor results were observed for HR (see Figure 3).

During informal interview reasons for not positively rating all max biometric response images as memorable or important at time included lack of novelty, e.g. regular lunch date with same group of collegues, subjects stated that audio for such events might have helped them recollect the event and hence change their ratings. Additionally events considered mundane by the subject, such as cooking dinner, while receiving max biometric response, received low ratings in the questionnaire. While we have no way of knowing what would cause such events to receive a max biometric response, we postulate that it may be due to the thought process of the subject at the time, which is now not recalled or some brief shock factor such as a sudden noise for example. Generally speaking we found that SenseCam images with max biometric response, which did not focus on interaction with other people, were not interesting to view. Exceptions here were images depicting a novel vacation location for example. Image processing techniques could potentially be exploited to further enhance the set of SenseCam images presented to a subject for browsing by down-weighting the potential selection of SenseCam images where people are not present and by up-weighting novel locations. The lifelog archives we have created contain people present information and geo-location information, as discussed in Section 2, which would enable such enhancements.

Strong correlation was not observed between GSR or HR levels and memorabelness, perceived importance of computer and mobile activity events and desire to view (Figure 2). However, correlation between ST levels and how interesting the subject now considered events was observed (see Figure 2). Additionally, subjects would or might want to retrieve (yes and not sure in Figure 4) 64% of max ST images, this compares with 40% of average ST images and 40% of

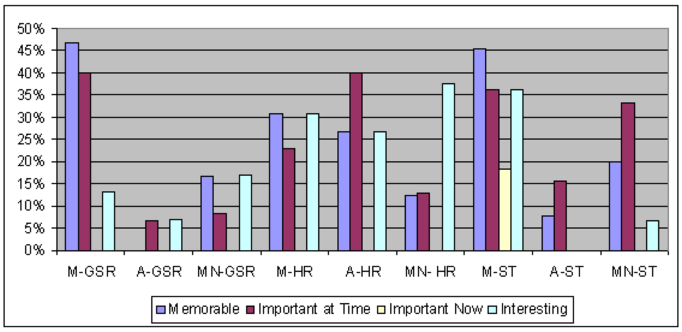


**Figure 1: Percentage of SenseCam events for max (M), average (A) and min (MN) GSR, HR and ST the subjects considered memorable, important at the time, important now, and are interested in viewing now.**

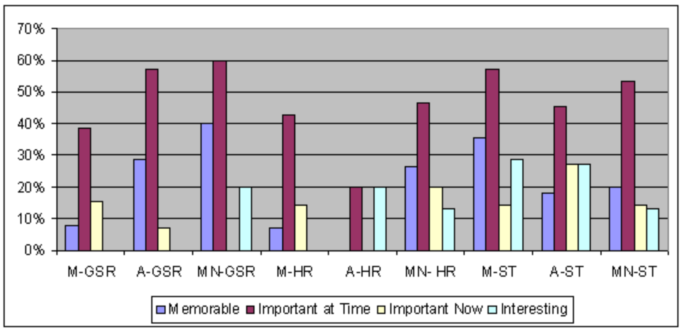


**Figure 2: Percentage of Computer and SMS events for max (M), average (A) and min (MN) GSR, HR and ST the subjects considered memorable, important at the time, important now, and are interested in viewing now.**

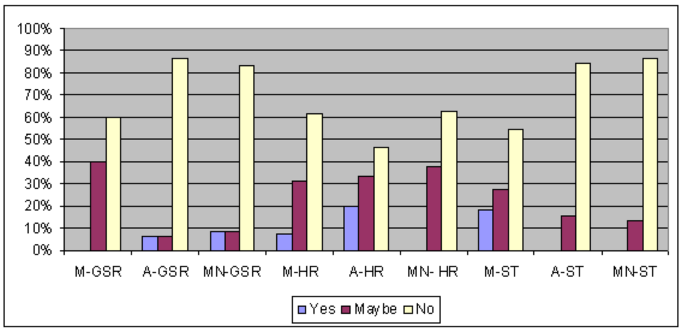


**Figure 3: Percentage of SenseCam events for max (M), average (A) and min (MN) GSR, HR and ST the subjects would (Yes), might (Maybe) and would not (No) want to view again.**

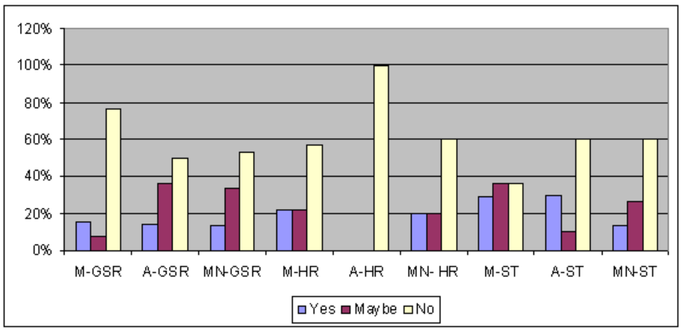


**Figure 4: Percentage of Computer and SMS events for max (M), average (A) and min (MN) GSR, HR and ST the subjects would (Yes), might (Maybe) and would not (No) want to view again.**

min ST images. No such correlation was observed for GSR or HR levels (Figure 4).

Informal interviews with subjects revealed some explanations for subjects' ratings of max biometric computer events. Completing a tedious user experiment stood out as memorable to subjects due to being particularly mundane, however subjects did not want to see these items again. Other events, such as emails containing information which was very important to subjects at the time of capture were very memorable, but as this information was no longer relevent to subjects they had no desire to view it again. Investigation of low ratings of instant messaging (IM) activity items with a max biometric response, revealed that subjects could not recall the content of IM which if available may have changed their ratings of these items. This result suggests that capturing additional data of items with max biometric values would be useful in further experimentation.

## 3.4 Concluding Remarks

The results of this experiment provide some perliminary support for the use of biometric response (especially ST) at time of item capture in extracting interesting items from large multimedia lifelog collections. We are interested in establishing if the results presented in this section can be improved using alternate approaches. We conducted similar experiments to the one described in this section to select max, average and min events, namely: 1) removing all items/events from the possible items to present to a subject in the max, average and min categories which occurred in more than one of these categories; 2) presenting items to subjects which occurred near in time to the noted max, average or min biometric responses. Result analysis is still on-going. Future work will explore combining GSR, HR and ST levels in the extraction of interesting events.

# 4. EXPERIMENT 2: RE-RANKING RESULT LISTS USING BIOMETRIC RESPONSE

In this section we describe the setup of a study to examine the utility of GSR and HR biometric data at original time of the computer item or SMS access, in re-ranking the output of a user query driven IR result list. For this investigation we used the 1 month collections of computer activity and SMS data annotated with biometric data from subjects' lifelogs. The results of this initial investigation do not support the use of GSR and HR biometric response at time of item creation/access for re-ranking lifelog query result lists; however interesting observations were made, as discussed later. We begin this section with a description of the experiment approach taken, and follow with an analysis of the results obtained.

## 4.1 Experiment Procedure

### 4.1.1 Test Case Generation

If lifelogs are to be recorded and accessed over an extended period it is important that users are able to reliably retrieve content recorded in the distant past. It is clear that a user is likely to remember a significant amount of content and context data soon after an event occurred, however with time memory fades and it is anticipated that less will be remembered a substantial delay after the event occurred [10]. In this experiment we wished to mimic the 'real' re-finding requirements of individuals, and details they are likely to recall about required items as closely as possible. As such, in generating the test cases for this experiment the following query generation technique was used:

- After 8 months lifelog collection build up (5 months after the one month biometric data capture period) subjects were required to list lifelog retrieval tasks they might want to perform in the future. This list was extended by the subjects consulting their lifelogs to determine additional items they might want to retrieve in the future. Typical test cases generated in this manner were: 'show me all documents I created associated with conference X; show me the SMS message my friend Sarah sent me regarding the location in Paris we were to meet in'.
- Subjects then entered their list of task descriptions along with keywords and remembered context, e.g. word in file name; extension type; month; day of week; weekday or weekend; morning, afternoon, evening or night; light status, into a provided form.

50 test cases were generated by each subject using this technique. Of these test cases subject 1 had 22 tasks containing items which occurred during the biometric data capture period, subject 2 had 8 and subject 3 had 36. These subsets of the generated test cases were used in this experiment.

### 4.1.2 Result Set Generation

To test the performace of our re-ranking approach a set of relevant computer/SMS items is required. Pooling is a commonly used approch in IR for generating lists of relevant items for a query. With pooling the query is passed into different IR algorithms and the results from each IR algorithm presented to subjects for relevance rating. We created our pooled result lists by entering various combinations of remembered content (keywords) and context into the vector space model (VSM) [22] and BM25 [20]. VSM and BM25 are good standard IR algorithms. They rank a set of documents in reponse to a user query using different term weighting approaches. The query types entered into VSM and BM25 were:

1. Content only.
2. Context only, this incorporated the following fields: word in file name; extension type; month; day of week; weekday or weekend; morning, afternoon, evening or night; light status.
3. Content + extension type.
4. Content + context.

The Lucene implementation of the vector space model and an in-house developed implementation of BM25 for Lucene were used to process these queries. Query type three and four are straightforward concatenations of the content data from query one with various types of the context data from query two.

The results from each of the 8 IR techniques were pooled and presented to subjects for relevance judgment. That is, for each test case the subjects rated each of the retrieved items as relevant or irrelevant. These judged sets were used for determining the utility of our techniques.

### 4.1.3 Investigation

We wished to investigate if biometric response (specifically GSR and HR) at time of item access/creation could be used to re-rank the output of an IR in response to a user query. To do this we investigated adding a query independent biometric score (static score) to items queried content and items queried content + context based retrieval score. The queried context fields used were: word in file name; extension type; month; day of week; weekday or weekend; morning, afternoon, evening or night; light status. While VSM was found to enrich the pooled result lists generated in Section 4.1.2, comparison showed BM25 to perform better. Hence BM25 was used to obtain queried content and queried content + context retrieval scores in this experiment. Static biometric scores were obtained by dividing the GSR associated with an item by the energy expenditure at that time, this value was multiplied by an empirically determined constant. The maximum static biometric score obtained in this manner across all accesses to the item was chosen as the static score value to add to the BM25 score. We also examined static biometric scores created by dividing HR by energy expenditure, again this value was multiplied by an empirically determined constant and the maximum score obtained for the item chosen to add as a static score to the BM25 item score.

For this investigation the following IR algorithms were investigated:

1. Content only
2. Content + GSR static score
3. Content + HR static score
4. Content + Context
5. Content + Context + GSR static score
6. Content + Context + HR static score

Query type one represents a good current standard approach for retrieval using search engines. Results generated from this content only query were used as a baseline.

The goal of our experiment is to retrieve the correct file(s) at top rank for a given query task. To investigate the usefulness of our static biometric scoring approaches in the retrieval process, the biometric data capture month of each subject's lifelog was queried with their set of queries using the 6 IR algorithms. The queries were processed using an in-house developed version of BM25 for Lucene. To obtain a relevance score for an item the relevance scores obtained

Table 1: **Average precision, precision @ 5 and @ 10 documents and R-precision results for 3 subjects.**

| IR Technique | AveP | P@5 | P@10 | Rprec |
|---|---|---|---|---|
| Content Only | 0.268366667 | 0.235566667 | 0.187933333 | 0.255233333 |
| Content Only+GSR | 0.248433333 | 0.2203 | 0.177166667 | 0.223 |
| Content Only+HR | 0.237566667 | 0.213966667 | 0.1708 | 0.2141 |
| Content+Context | 0.3025 | 0.257666667 | 0.208466667 | 0.261866667 |
| Content+Context+GSR | 0.271566667 | 0.2349 | 0.195633333 | 0.225466667 |
| Content+Context+HR | 0.256766667 | 0.2402 | 0.184833333 | 0.206633333 |

for the items content and each of the items context types were summed, and the static biometric score added to this. In each case the rank of relevant items in the result set was noted. The next section discusses the results of this investigation.

## 4.2 Experiment Results

TREC_EVAL [25], a freely available IR evaluation tool was used to calculate the precision at 5 and 10 documents, average precision, and R-precision of retrieved events. R-precision is the recall result, it measures precision after the total number of relevent items for a query have been retrieved (e.g. if there were 9 relevant items for a query, precision would be calculated after 9 documents had been retrieved). Results obtained are shown in Table 1. As can be seen from the results adding a biometric static score to content or content + context IR scores does not prove useful. Substantial drop in results is noted across average precision, R-precision and precision at 5 and 10 documents with the addition of either static GSR or HR scores. These preliminary results suggest that biometric response at time of item creation/access is not useful for re-ranking lifelog text-based collections. We discovered that, for a given task, biometric response of relevant retrieved items serves as a good indicator of a user's engagement with a computer item at time of creation/access, however this does not necessarily mean that the item will be useful in the future. For example, an activity such as registering for a conference may cause a marked biometric response at the time, but this does not mean that an individual will find this registration form useful in the future, nor are they likely to consider it important if they perform a future query for information relating to the conference. Similarly, twittering about the frustrating Java coding task you are currently engaged in may cause a marked biometric response, but this twitter will not be considered relevant in a future query on how to code in 'Java'.

## 4.3 Concluding Remarks

Based on the results of this perliminary experiment, GSR and HR levels at time of item access do not seem useful in re-ranking user query driven IR result lists. The experiment in this section was conducted prior to Experiment 1. In light of the correlation between ST and importance and desire to view items observed in Experiment 1, ST may prove to be a more useful static biometric score. Experimentation to test this hypothesis is currently in progress.

# 5. CONCLUSIONS

In this paper we set out to explore the role of biometric response in lifelog item/event retrieval. We investigated whether items coincident with maximum observed biometric galvanic skin response (GSR) and heart rate (HR) and with minimum observed skin temperature (ST) readings were more important to subjects and whether this would mean they were most useful for presentation to subjects when browsing their personal information archives. From this preliminary study correlation between GSR and ST levels and SenseCam event importance was observed. The SenseCam event selection results are important since ability to extract interesting events from vast SenseCam collections is challenging but important if these archives are to have long-term use. From this preliminary study GSR and ST appear potential sources of evidence for selection of SenseCam images. Combining GSR and ST responses may prove to further improve this performance. Correlation between ST response and computer item and SMS importance was also observed. While these results are promising, it is acknowledged that this study was conducted on a limited number of subjects over a short period of time. Investigation using more participants over a longer timeframe is required to further test our conclusions.

Our second investigation looked at the utility of using GSR or HR level at time of item creation/access to boost results in a query driven IR result list. Our preliminary results suggest that biometric response at original time of item access, while indicating the current importance of, or engagement with an item does not reflect the future likelihood of this item being relevant to a given query. We are currently investigating if ST may prove more useful.

Developments in technology are enabling individuals to store increasing amounts of digital data pertaining to their lives. As these personal archives grow ever larger, reliable ways to help individuals locate interesting items from these multimedia lifelogs becomes increasingly important. The results of these experiments provide perliminary indication that biometric response, in particular ST, may serve as a useful tool for extracting interesting items from long-term multimedia lifelogs. Additionally, the observed correlation between biometric response and current engagement found in these studies suggests that biometric response could be useful in 'live' applications such as 'on-line recommender systems (e.g. shopping websites), web search systems, or archive search systems (e.g. searching through picture archives or document archives) where the subject's search could be guided by their biometric response to items suggested / viewed at each iteration of the recommendation / search process.

# 6. ACKNOWLEDGMENTS

We would like to acknowledge the test subjects for their contribution to this study.

This work is funded by a grant under the Science Foundation Ireland Research Frontiers Programme 2006, and supported by Microsoft Research under grant 2007-056.

## 7. REFERENCES


[1] Apache Lucene. Overview, Copyright ©2006 The Apache Software Foundation. http://lucene.apache.org/java/docs/.

[2] G. Bell. Challenges in Using Lifetime Personal Information Stores based on MyLifeBits. Alpbach Forum, 2004.

[3] BodyMedia - http://www.bodymedia.com.

[4] A. Damsaio. *Descartes' Error Emotion, Reason and the Human Brain*. Bard, 1994.

[5] B. Ferry, B. Roozendaal, and J. McGaugh. Basolaterail amygdala noradrenergic influences on memory storage are mediated by an interaction between beta- and alpha1-adrenoceptors. *Journal of Neuroscience*, 19:5119–5123, 1999.

[6] M. Gazzaniga, R. Ivry, and G. Mangun. *Cognitive Neuroscience (Second Edition)*. Norton, 2002.

[7] J. Gemmell, G. Bell, and R. Lueder. MyLifeBits: A Personal Database for Everything. *Communications of the ACM. Personal Information Management*, 49:88–95, 2006.

[8] C. Gurrin, A. Smeaton, D. Byrne, N. O'Hare, G. Jones, and N. O'Connor. An Examination of a Large Visual Lifelog. In *4th Asia Information Retrieval Symposium (AIRS)*, pages 537–542, 2008.

[9] A. Joki, J. Burke, and D. Estrin. Campaignr - a framework for participatory data collection on mobile phones. Technical report, 2007.

[10] L. Kelly, Y. Chen, M. Fuller, and G. Jones. A study of remembered context for information access from personal digital archives. In *2nd International Symposium on Information Interaction in Context (IIiX)*, pages 44–50, 2008.

[11] V. Kostakos, E. O'Neill, and S. Jones. Social networking 2.0. In *Extended abstracts of the conference on Human factors in computing systems*, pages 3555–3560, 2008.

[12] L. Kelly and G.J.F. Jones. Venturing into the Labyrinth: the Information Retrieval Challenge of Human Digital Memories. In *Proceedings of the Workshop on Supporting Human Memory with Interactive Systems, at HCI 2007: The 21st British HCI Group Annual Conference.*

[13] P. Lang. The emotion probe: Studies of motivation and attention. *American Psychologist*, 50(5):372–385, 1995.

[14] B. Lavelle, D. Byrne, C. Gurrin, A. Smeaton, and G. Jones. Bluetooth Familiarity: Methods of Calculation,Applications and Limitations. In *MIRW 2007 - Mobile Interaction with the Real World, Workshop at the MobileHCI07: 9th International Conference on Human Computer Interaction with Mobile Devices and Services*, pages 55–58, 2007.

[15] LogExport - http://tinyhack.com/freewarelist/s603rd/2007/03/02/logexport/.

[16] J. McGaugh. *Strong memories are made of this Memory and Emotion: The Making of Lasting Memories*. Columbia University Press, 2003.

[17] MSR MyLifeBits Project - http://research.microsoft.com/barc/mediapresence/MyLifeBits.aspx.

[18] Nokia N95 mobile phone - http://www.forum.nokia.com/devices/N95.

[19] Polar Heart Rate Monitor - http://www.polarusa.com/.

[20] S. Robertson and K. S. Jones. Simple, proven approaches to text retrieval. Technical Report 356, University of Cambridge, 1994.

[21] R. Sakamoto, A. Nozawa, H. Tanaka, T. Mizuno, and H. Ide. Evaluation of the driver's temporary arousal level by facial skin thermogram-effect of surrounding temperature and wind on the thermogram. *IEEJ Trans. EIS*, 126(804-9), 2006.

[22] G. Salton and C. Buckley. Term-weighting approaches in automatic text retrieval. In *Information Processing and Management*, pages 513–523, 1988.

[23] MSR SenseCam - http://research.microsoft.com/sendev/projects/sensecam.

[24] Slife Labs - http://www.slifelabs.com.

[25] trec_eval - http://trec.nist.gov/trec_eval/.