Biometric Response as a Source of Query Independent Scoring in Lifelog Retrieval

Liadh Kelly and Gareth J.F. Jones

Centre for Digital Video Processing,
Dublin City University, Dublin 9, Ireland
{lkelly,gjones}@computing.dcu.ie

Abstract. Personal lifelog archives contain digital records captured from an individual's daily life, e.g. emails, web pages downloaded and SMSs sent or received. While capturing this information is becoming increasingly easy, subsequently locating relevant items in response to user queries from within these archives is a significant challenge. This paper presents a novel query independent static biometric scoring approach for re-ranking result lists retrieved from a lifelog using a BM25 model for content and content + context data. For this study we explored the utility of galvanic skin response (GSR) and skin temperature (ST) associated with past experience of items as a measure of potential future significance of items. Results obtained indicate that our static scoring techniques are useful in re-ranking retrieved result lists.

Key words: Lifelog retrieval, affective response, query independent weights
.

1 Introduction

Advances in digital technologies mean a wealth of personal information is now becoming available in digital format. This information can be gathered together and stored in a personal lifelog (PL) [3]. Lifelog archives can contain everything from items read, written, or downloaded; to footage from life experiences, e.g. photographs taken, 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 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. One potential source of useful information is the user's biometric response associated with previous experience with an item. In this study we explore two biometric responses associated with items, namely galvanic skin response (GSR) and skin temperature (ST).

Previous work has shown an individual's biometric response to be related to their overall arousal levels [11]. Significant or important events tend to raise an individual's arousal level, causing a measurable biometric response [12]. Events that can be recalled clearly in the future are often those which were important or emotional in our lives [7]. 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 [6]. 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 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 [7]. Arousal response can also be observed through ST. 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 ST [15]. 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 [4] which can continuously record the wearer's GSR and ST.

We propose that lifelog items which are important to an individual at the time they were experienced 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 accessing these items. Thus recording GSR and ST as part of a lifelog may enable us to identify important items which would be most important in the context of this paper, to a given future information searching task. In particular we hypothesize that adding a query independent boost (static score) to important items in lifelog IR result lists, where important items are detected based on recorded GSR or ST levels associated with past accesses to the items, may improve retrieval performance.

In this paper we report our findings to date which may guide future research in this area. We describe our study to investigate the utility of biometric response in re-ranking traditional information retrieval result lists. We find positive results for a technique for adding static biometric scores to the results of content or content+context result lists obtained using a BM25 retrieval model.

The next section discusses related work and highlights the contributions of this paper. Section 3 describes the test-set gathered for this study. Section 4 presents our experimental set-up and results are discussed in Section 5. We conclude the paper with a discussion of findings and directions for future work.

2 Towards Static Biometric Scores

While observed biometric response has been used to detect tasks or items in different test sets which are of current relevance or importance to the individual, for example movie scene selection [16] and elicitation of topical relevance in multimedia systems [2], to our knowledge previous research has not investigated the exploitation of observed biometric response as an implicit indicator of future item importance. This we believe is an important previously unexploited opportunity to gain passive feedback from subjects for improving the retrieval performance of future searches in both lifelogging and other domains. In a preliminary study we found correlation between lifelog items coincident with maximum observed GSR and with minimum observed ST at the time of item creation/access and current importance of the items [10] . This finding motivates our current study to investigate the utility of adding a static biometric relevance score generated using a BM25 model for content or content+context retrieval.

There are many examples of the use of various types of static scores to boost user query driven scores in different domains. Examples here include the well known PageRank which uses the webs link structure to create static scores for web pages [13], using web page features such as document length and anchor text as static scores [14], and using links created between computer files to infer static file importance scores [17].

Various approaches can be used for integrating static scores with query dependent scores. We explore using a linear combination of the query dependent score and static biometric score. We also investigate various approaches for transforming the biometric response into a static score. In particular raw biometric scores and various nonlinear transformations of the biometric readings are explored. A particularly promising technique is presented in [5] where a sigmoid functional form is used to transform PageRank, link indegree, ClickDistance and URL length into static scores. This technique forms part of our investigation.

3 Test-set

In order to explore our hypothesis, a suitable test-set must be available. As part of our ongoing work on PLs we are gathering long term multimedia lifelog collections, stored locally on individuals PCs, from a small group of subjects [8]. For the current investigation we augmented these lifelogs for 3 postgraduate students within our research group (1 male, 2 females; from Asian and Caucasian ethnic groups), for a 1 month period, with capture of their GSR and ST data.

GSR, ST and energy expenditure were collected using a BodyMedia armband [4] worn on the upper arm. Based on results from initial calibration experiments, GSR data was capture once per second and ST data once every ten seconds[1].

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 associated with meaningful variations in arousal levels, as opposed to physical activity. Energy expenditure (sampled once per minute) correlates well with physical activity levels. Thus measured energy expenditure can be used to differentiate between high GSR and low ST biometric data levels, resulting from physical activity and those arising from events experienced from the environment.

In addition to the biometric data, our 1 month experimental lifelogs contained data of computer activity and SMSs sent and received. See [8] for full details on data capture. For this experiment we used the content of SMSs and computer

[1] Due to an error in settings ST was sampled once per minute for Subject 3.

items (e.g. word document) created/accessed annotated with context types[2]: word in file name; extension type; month; day of week; weekday or weekend; is beginning of week; is mid-week; is end week; is morning; is afternoon; is evening; is night. 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).

4 Experiment Procedure

In this section we describe the setup of our study to examine the utility of GSR and ST biometric data at the time of previous computer item or SMS access, in re-ranking the output of a user query driven IR result list. We begin this section by describing our test case and result set generation approach, and follow with details of our investigation and static scoring approaches.

4.1 Test Case Generation

If PLs 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. 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 [9]. Query generation in the PL domain is challenging. 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. In generating the test cases for this experiment the following approach was used to generate 50 queries per subject:

- After 8 months lifelog collection build up (5 months after the one month biometric data capture period) subjects listed lifelog retrieval tasks they might want to perform in the future. Typical test cases generated in this manner were: 'show me documents I created associated with conference X'.
- Subjects then entered their list of task descriptions along with keywords and remembered context, e.g. extension type, into a provided form.

4.2 Result Set Generation

Pooled result lists were created by entering content (keywords) only, context only, content+extension type, and content+context query types into two good standard retrieval systems, namely the vector space model (VSM) and BM25, to retrieve as many relevant items from subjects' collections as possible. The BM25 k and b parameters tuned to 1.5 and 1 respectively using the full set of each user's queries. The Lucene implementation of the VSM and an in-house developed implementation of BM25 for Lucene were used to process these queries. Queries combining content and context are straightforward concatenations of the content data score with the individual context types scores. The results from each of the

[2] In experiment section context data refers to all these context types.

8 IR techniques were pooled and presented to subjects for relevance judgment (i.e. 0 = irrelevent; 1 = relevant). These judged sets were used for determining the utility of our techniques.

4.3 Investigation

We investigated if GSR and ST at the time of item experience could be used to re-rank the output of IR in response to a user query. Queries used for this investigation were those contained in the subset of the 50 test cases generated for each subject which contained items occurring during the biometric capture month. Subject 1 had 22 such tasks, Subject 2 had 8 and Subject 3 had 36.

While VSM was found to enrich the pooled result lists generated in Section 4.2, comparison showed BM25 to perform better in retrieval. Hence our in-house developed version of the BM25 system for Lucene was used to obtain queried content and queried content+context retrieval scores in this experiment. For content+context querying, the relevance scores obtained for the items content and each of the item's context types were summed. The weight (w) assigned to each field and BM25 k and b parameters were tuned using the full set of the 3 subjects biometric month test cases. Only the top 1000 results were taken in each case for efficiency, without a serious degradation in performance. Static biometric scores were added to the content and content+context scores (techniques used to obtain static biometric scores are described in Section 4.4). In each case the rank of the relevant items in the result set was noted. For content only retrieval 4116 items were retrieved for Subject 1 (relevant: 90, rel_ret: 40), 84 items were retrieved for Subject 2 (relevant: 16, rel_ret: 0), and 16768 items were retrieved for Subject 3 (relevant: 556, rel_ret: 480). For content+context retrieval 11912 items were retrieved for subject 1 (relevant: 90, rel_ret: 90), 3385 items were retrieved for subject 2 (relevant: 16, rel_ret: 16), and 28132 items were retrieved for subject 3 (relevant: 557, rel_ret: 530).

4.4 Static Relevance Scores

Each retrieved item for content only retrieval was annotated with the maximum observed GSR and associated energy expenditure (engGSR) and with the minimum observed ST and associated energy expenditure (engST), across all accesses to the item. Items with no associated biometric readings, due to biometric recording devices being removed for data downloading purposes, subjects' need for mental break from wearing of devices, etc, were assigned default biometric values. The default value used was the average of the GSR, engGSR, ST and engST readings associated with retrieved items.

Increases in physical activity (detected through increases in energy expenditure) cause GSR levels to increase and ST levels to decrease. To discern changes in GSR caused by changes in arousal level as opposed to changes in physical activity, we also tagged items with GSR divided by engGSR. As stated in Section 1 the lower the ST level the greater the arousal level, hence the inverse of ST and the inverse of ST divided by engST levels (to account for changes in physical activity) associated with retrieved items were also tagged to items. The GSR, inverse ST, GSR divided by engGSR, and inverse ST divided by engST values associated with retrieved items were normalised using min-max normalisation.

To allow for investigation of the approach, mentioned in Section 2, which calculates static relevance scores for features where lower values indicate greater importance, we also normalised the ST values and ST values multiplied by energy expenditure using min-max normalisation and tagged these values to items.

The same process was also applied to tag items retrieved from content+context retrieval with GSR and ST levels. The following approaches for calculating static relevance scores using the normalised biometric data tags were investigated:

$$STbase = w \cdot \frac{\frac{1}{ST}}{engST} \tag{1}$$

$$logST = w \cdot log(\frac{1}{ST}) \tag{2}$$

$$logSTdivEng = w \cdot log(\frac{\frac{1}{ST}}{engST}) \tag{3}$$

$$sigmST = w \cdot \frac{s^a}{k^a + s^a},\ where\ s = \frac{1}{ST} \tag{4}$$

$$sigmSTdivEng = w \cdot \frac{s^a}{k^a + s^a},\ where\ s = \frac{\frac{1}{ST}}{engST} \tag{5}$$

$$sigmIncST = w \cdot \frac{k^a}{k^a + s^a},\ where\ s = ST \tag{6}$$

$$sigmIncSTmultEng = w \cdot \frac{k^a}{k^a + s^a},\ where\ s = ST \times engST \tag{7}$$

$$GSRbase = w \cdot \frac{GSR}{engGSR} \tag{8}$$

$$logGSR = w \cdot log(GSR) \tag{9}$$

$$logGSRdivEng = w \cdot log(\frac{GSR}{engGSR}) \tag{10}$$

$$sigmGSR = w \cdot \frac{s^a}{k^a + s^a},\ where\ s = GSR \tag{11}$$

$$sigmGSRdivEng = w \cdot \frac{s^a}{k^a + s^a},\ where\ s = \frac{GSR}{engGSR} \tag{12}$$

Equations 1 and 8 are our baseline static scoring approaches, used to examine the effect of the raw ST and GSR values divided by energy expenditure on re-ranking result lists. The remaining equations investigate the use of non-linear transformations of the biometric score. Equations 2, 3, 9, 10 examine the effect of using logs of ST and GSR. The performance of our biometric scores using the transformation approach presented in [5] is examined with Equations 4, 5, 11, 12. This approach is used to generate static relevance scores for features where higher values indicate greater importance. An approach for calculating static relevance scores for features where lower values indicate greater importance is also provided in [5]. This techniques performance using our ST data is investigated with Equations 6 and 7. The effect of accounting for energy expenditure is investigated in Equations 1, 3, 5, 7, 8, 10, 12. Following parameter tuning using the full set of the 3 subjects' biometric month test cases, the static score's weight of importance (w) and parameters k and a were set for each equation.

The static scoring techniques presented in this section are added to content and content+context relevance scores generated using BM25 model, described in Section 4.3. The next section discusses results obtained using these approaches.

5 Experiment Results and Analysis

Average precision (AveP), P@5 and P@10 were investigated. P@5 and P@10 show how effective our techniques were at moving relevant items towards the top of the result lists. Table 1 shows the percentage improvement over the content only baseline for content+static_score retrieval averaged over Subjects 1 and 3 (relevant items were not retrieved for Subject 2 using content only retrieval). Percentage improvement for content+context+static_score retrieval over the content+context baseline, averaged over all 3 subjects, are also presented in Table 1. Table 2 presents the individual breakdown of results for each subject. Overall results suggest that adding either a GSR or ST static score to content or content+context IR scores is useful for re-ranking PL text-based collections. In particular, the use of ST as a static score yields the greatest overall improvement in performance. In this section we analyse the results obtained and suggest a general function for calculating query independent biometric scores for re-ranking BM25 model generated content and content+context PL result lists.

5.1 Overall Static Score Performance

Considering both content and content+context retrieval the addition of a static score using *sigmSTdivEng* resulted in the greatest percentage improvement from the content and content+context baselines. 0%, 5% and 28% improvement in average precision, P@5 and P@10 respectively were observed for content only retrieval using this technique. While content+context retrieval yielded 4%, 34% and 6% improvement for average precision, P@5 and P@10 respectively. The lower percentage in improvement for content+static_ST_score may be explained by the lack of retrieved items for content only IR for Subject 2 who benefited the most from the addition of a static ST score (see Table 2).

The superior performance of *sigmSTdivEng* to *STbase* and to the approaches which calculate logs of the ST scores is consistent with the findings noted in [5] where this transformation was used with the greatest success in calculating

Table 1. Average percentage improvement, rounded to nearest whole number, by adding a static score (staticS) to the content (C) and content+context (CC) baselines.

Static Technique	**C+staticS**			**CC+staticS**		
	AveP	**P@5**	**P@10**	**AveP**	**P@5**	**P@10**
STbase	-2%	-3%	-4%	3%	35%	-1%
logST	0%	1%	1%	1%	3%	6%
logSTdivEng	0%	1%	0%	4%	34%	7%
sigmST	0%	6%	34%	0%	3%	6%
sigmSTdivEng	0%	5%	28%	4%	34%	6%
sigmIncST	1%	2%	2%	1%	1%	8%
sigmIncSTmultEng	0%	-1%	-2%	-1%	0%	6%
GSRbase	0%	-1%	0%	-2%	32%	6%
logGSR	1%	3%	2%	-2%	35%	7%
logGSRdivEng	0%	1%	2%	-1%	34%	8%
sigmGSR	1%	3%	2%	-2%	35%	8%
sigmGSRdivEng	0%	1%	2%	-1%	34%	8%

weights for static features of web pages. When using this transformation to calculate query independent ST scores we took the inverse of ST, since lower ST levels indicate greater importance. However [5] also presents a transformation for calculating static scores for situations where lower static values indicate greater importance. This technique did not perform as well overall on our test-set, see results *sigmIncST* and *sigmIncSTmultEng* in Table 1.

While adding GSR static scores also improved retrieval performance, albeit not to the same extent as the addition of ST static scores, no clear best approach for adding GSR static scores was detected. Greatest improvement observed for content only retrieval by the addition of a static GSR score was 1%, 3% and 2% for AveP, P@5 and P@10 respectively using *logGSR* or *sigmGSR*, as shown in Table 1. In all cases average precision was decreased by the addition of a static GSR score, while P@5 increased by 32-35% and P@10 increased by 6-8% for content+context retrieval. Additionally unlike ST static scores, the performance of GSR static scores were not greatly altered overall by factoring in energy expenditure (see Table 1).

5.2 Performance Across Individual Subjects

Exploring the individual results of each subject (see Table 2), we find that results for Subject 2 were greatly improved by the addition of a static score to the base content+context score. This improvement was observed equally for *logSTdivEng* and *sigmSTdivEng*, and to a lesser extent for *GSRbase*, *logGSRdivEng* and *sigmGSRdivEng*, with a 100% improvement in precision @5 in all cases.

Subject 3 benefited the least from the introduction of a static ST score, with 1% increase in P@5 being observed. Similar to Subject 2, greatest improvement was noted using the *logSTdivEng* and *sigmSTdivEng* transformations to calculate static ST scores. However, no variation in performance was observed between the five static GSR scoring techniques for this subject.

Table 2. Subjects' percentage improvement, rounded to nearest whole number, for Average precision (AveP), P@5 and P@10 by adding a static score to the content and content+context baselines.

Static Technique	Subject 1			Subject 2			Subject 3		
	AveP	P@5	P@10	AveP	P@5	P@10	AveP	P@5	P@10
	Content + static score								
STbase	0%	-4%	-3%	-	-	-	-4%	-1%	-6%
logST	2%	0%	0%	-	-	-	0%	1%	-1%
logSTdivEng	0%	0%	0%	-	-	-	0%	1%	0%
sigmST	0%	12%	69%	-	-	-	0%	1%	-1%
sigmSTdivEng	-1%	9%	56%	-	-	-	0%	1%	0%
sigmIncST	1%	4%	3%	-	-	-	0%	0%	0%
sigmIncSTmultEng	1%	-4%	-3%	-	-	-	0%	1%	0%
GSRbase	-1%	-4%	0%	-	-	-	0%	1%	0%
logGSR	2%	4%	3%	-	-	-	0%	1%	0%
logGSRdivEng	1%	0%	3%	-	-	-	0%	1%	0%
sigmGSR	2%	4%	3%	-	-	-	0%	1%	0%
sigmGSRdivEng	1%	0%	3%	-	-	-	0%	1%	0%
	Content + Context + static score								
STbase	1%	4%	-3%	9%	100%	0%	-1%	1%	-1%
logST	2%	8%	0%	1%	0%	20%	0%	1%	-1%
logSTdivEng	0%	0%	0%	11%	100%	20%	0%	1%	0%
sigmST	0%	8%	0%	1%	0%	20%	0%	1%	-1%
sigmSTdivEng	0%	0%	-3%	11%	100%	20%	0%	1%	0%
sigmIncST	1%	4%	3%	1%	0%	20%	0%	0%	0%
sigmIncSTmultEng	1%	0%	-3%	-4%	0%	20%	0%	1%	0%
GSRbase	-2%	-4%	-3%	-4%	100%	20%	0%	1%	0%
logGSR	1%	4%	0%	-7%	100%	20%	0%	1%	0%
logGSRdivEng	0%	0%	3%	-4%	100%	20%	0%	1%	0%
sigmGSR	1%	4%	3%	-7%	100%	20%	0%	1%	0%
sigmGSRdivEng	0%	0%	3%	-4%	100%	20%	0%	1%	0%

Subject 1 benefited more than Subject 3 from the addition of static scores to the base content and content+context retrieval scores. For Subject 1, in contrast to Subjects 2 and 3, greater performance was observed when we did not divide by energy expenditure while calculating ST static scores (see results for *logST* and *sigmST* in Table 2). Greatest improvement in performance was observed by the addition of ST static scores to the base content only score using *sigmST* for this subject (12% improvement for P@5 and 69% improvement for P@10). Of the static GSR scoring approaches *sigmGSR* proved most useful for this subject. On biometric data analysis we found that Subject 1 had unexplained periods of particularly high energy expenditure relative to the other subjects which caused energy expenditure to be less useful for this subject.

On analysing the minor improvement in results observed for Subject 3 we found this subject to have a higher average precision for content only retrieval (=0.4715) and for content+context retrieval (=0.4593), than Subject 1 (content only = 0.3317, content+context = 0.3421) and Subject 2 (content+context = 0.1005). The higher default AveP values for this subject may partially explain the minor improvements introduction of static scores made for this subject. Conversely Subject 2 with the lowest default average precision benefited the most from the introduction of static scores. Subject 3's ST BM25+static_score values might have been further affected by the fact that ST was only sampled once per minute for this subject (compared to once every 10 seconds for the other subjects), as discussed in Section 3. Finally the high percentage of content and content+context query results which were assigned the default ST and GSR scores might also have impacted on Subject 3's results. 58% of content+context and 58% of content only retrieved items for Subject 3 did not have ST and GSR values associated with them, and hence were assigned the default ST and GSR values. This compares with 35% of content+context retrieved results and 29% of content only retrieved items for Subject 1 and with 41% of content+context retrieved items for Subject 2.

5.3 Biometric Static Scoring Function

Overall static ST scores provided greatest improvement. *logSTdivEng* and *sigmSTdivEng* performed best for Subjects 2 and 3 where the engST range tagged to retrieved lifelog items was quite narrow. *sigmST* or *logST* worked better for Subject 1 whose data contained unusually high engST readings. Analyses of engST readings which were captured by the biometric device and tagged to items retrieved for content+context retrieval revealed the following for Subject 1 (rounded to 2 decimal places): average = 2.19; median = 1.45; max = 12.10; min = 1.16. For Subject 2: average = 1.13; median = 1.07; max = 3.77; min = 1.03. For Subject 3: average = 1.34; median = 1.23; max = 4.88; min = 1.09.

Median and min values for Subject 1, 2 and 3 are in line. However, Subject 1 has a much larger max engST reading, of 12.10, than Subjects 2 and 3. Subject 1's median and max values indicate that they had infrequent unexplained periods of unusually high energy expenditure, which would somewhat degrade the static ST biometric scoring performance observed when dividing ST by engST. Of greater consequence though perhaps is the fact that their average engST is also higher as a result of the unusually high energy expenditure readings. This higher average engST was the default engST value assigned to items in the result list with no recorded biometric data. In all probability this negatively affected the results in Table 2 for Subject 1 where engST was factored. This analysis, leads us to the following approach for calculating the static ST score to add to the base BM25 retrieval score:

if (median engST * 5) $\leq$ max engST : use *sigmST*,
else : use *sigmSTdivEng*.

5.4 Concluding Remarks

While this study was performed on a limited number of subjects, it provides preliminary support for the use of biometric static scores, in particular ST, to boost relevant retrieved items in lifelogs and supports investment of further research in this space. We are interested in examining the scalability of our approach using larger numbers of subjects and in determining if the results presented in this section can be improved using alternate approaches. In particular, given the findings from our analysis of energy expenditure values, future work will explore the use of alternate approaches for calculating the default values assigned to items missing biometric response, for example using median values instead of averages. The affect on performance when GSR and ST readings are combined will also be looked at. As well as the use of alternate biometric readings for calculating static scores, for example using heat flux and heart rate.

6 Conclusions

In this paper we set out to investigate the role of biometric response in lifelog item retrieval. We presented a novel approach for calculating static relevance scores to boost results in a query driven IR result list using an individual's biometric response at the original time of item access. Results obtained support the use of this approach. Greatest improvement in performance was found by the addition of a static skin temperature (ST) score. From these results a general function for calculating query independent ST scores was derived.

While these results are promising, it is acknowledged that this study was conducted on a limited number of subjects over a relatively short period of time. Further experiments with larger numbers of subjects are required to establish the scalability of the technique presented in this paper. However, due to the large psychological burden placed on subjects wearing the biometric devices for extended periods of time, and the difficulty in gaining participants willing to partake in experiments which log their personal data, this initial study formed a good means to establish if further research in this domain is warranted. Given the results presented in this paper we believe it is worth investing in further research in this space using larger collections of subjects.

Technological developments 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 required items from these lifelogs becomes increasingly important. The results of these experiments indicate that static biometric scores, in particular ST, serve as a useful tool for aiding extraction of important items from long-term lifelogs. Additionally, beyond the lifelogging domain, we envisage several possible applications of the technique presented in this paper both in the archive searching and recommendation spaces. Indeed in a future where biometric recording is prevalent, the same patterns of biometric response may be observed across individuals for the same items in shared archives (e.g. digital libaries, photo archives, retail websites), which might allow such items to be given query independent boosts for all users of the archive. Current research exploring development of less cumbersome biometric recording devices, for example research at MIT Media Lab, provides indication that reliable unobtrusive biometric devices embedded in individuals clothes or braclets for example will be widely available for use by such tools.

Acknowledgments. This work is funded by a grant under the Science Foundation Ireland Research Frontiers Programme 2006 Grant No: 06/RFP/CMS023.

References

1. Apache Lucene. http://lucene.apache.org/java/docs/.
2. I. Arapakis, I. Konstas, and J. Jose. Using Facial Expressions and Peripheral Physiological Signals as Implicit Indicators of Topical Relevance. In *Proceedings of the ACM International Conference on Multimedia*, pages 461–470, 2009.
3. G. Bell. Challenges in Using Lifetime Personal Information Stores based on MyLifeBits. Alpbach Forum, 2004.
4. BodyMedia - http://www.bodymedia.com.
5. N. Craswell, S. Robertson, H. Zaragoza, and M. Taylor. Relevance Weighting for Query Independent Evidence. In *Proceedings of the Annual ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 416–423, 2005.
6. B. Ferry, B. Roozendaal, and J. McGaugh. Basolateral Amygdala Noradrenergic Influences on Memory Storage Are Mediated by an Interaction between beta- and alpha1-Adrenoceptors. *Journal of Neuroscience*, 19(12):5119–5123, June 1999.
7. M. S. Gazzaniga, R. B. Ivry, and G. R. Mangun. *Cognitive Neuroscience (Second Edition)*. Norton, 2002.
8. L. Kelly, D. Byrne, and G. J. F. Jones. The role of places and spaces in lifelog retrieval. In *PIM 2009 - Proceedings of Personal Information Management, Workshop at ASIST 2009*, 2009.
9. L. Kelly, Y. Chen, M. Fuller, and G. J. F. 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.
10. L. Kelly and G. J. F. Jones. Examining the utility of affective response in search of personal lifelogs. In *5th Workshop on Emotion in HCI, British HCI Conference 2009*, 2009.
11. P. J. Lang. The emotion probe: Studies of motivation and attention. *American Psychologist*, 50(5):372–385, 1995.
12. J. McGaugh. *Strong memories are made of this Memory and Emotion: The Making of Lasting Memories*. Columbia University Press, 2003.
13. L. Page, S. Brin, R. Motwani, and T. Winograd. The PageRank Citation Ranking: Bringing Order to the Web. Technical report, January 1998.
14. M. Richardson, A. Prakash, and E. Brill. Beyond PageRank: Machine Learning for Static Ranking. In *Proceedings of the International World Wide Web Conference*, pages 707–715, 2006.
15. 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(7):804–809, 2006.
16. M. Soleymani, G. Chanel, J. J. Kierkels, and T. Pun. Affective ranking of movie scenes using physiological signals and content analysis. In *Proceedings of the 2nd ACM workshop on Multimedia Semantics*, pages 32–39, 2008.
17. C. A. N. Soules. *Using context to assist in personal file retrieval*. PhD thesis, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, USA, 2006.