AffectAura: An Intelligent System for Emotional Memory

Daniel McDuff[1,2], Amy Karlson[1], Ashish Kapoor[1], Asta Roseway[1], Mary Czerwinski[1]
[1]Microsoft Research, One Microsoft Way, Redmond, USA
[2]MIT Media Lab, 75 Amherst St, Cambridge, USA
djmcduff@mit.edu, {karlson, akapoor, astar, marycz}@microsoft.com

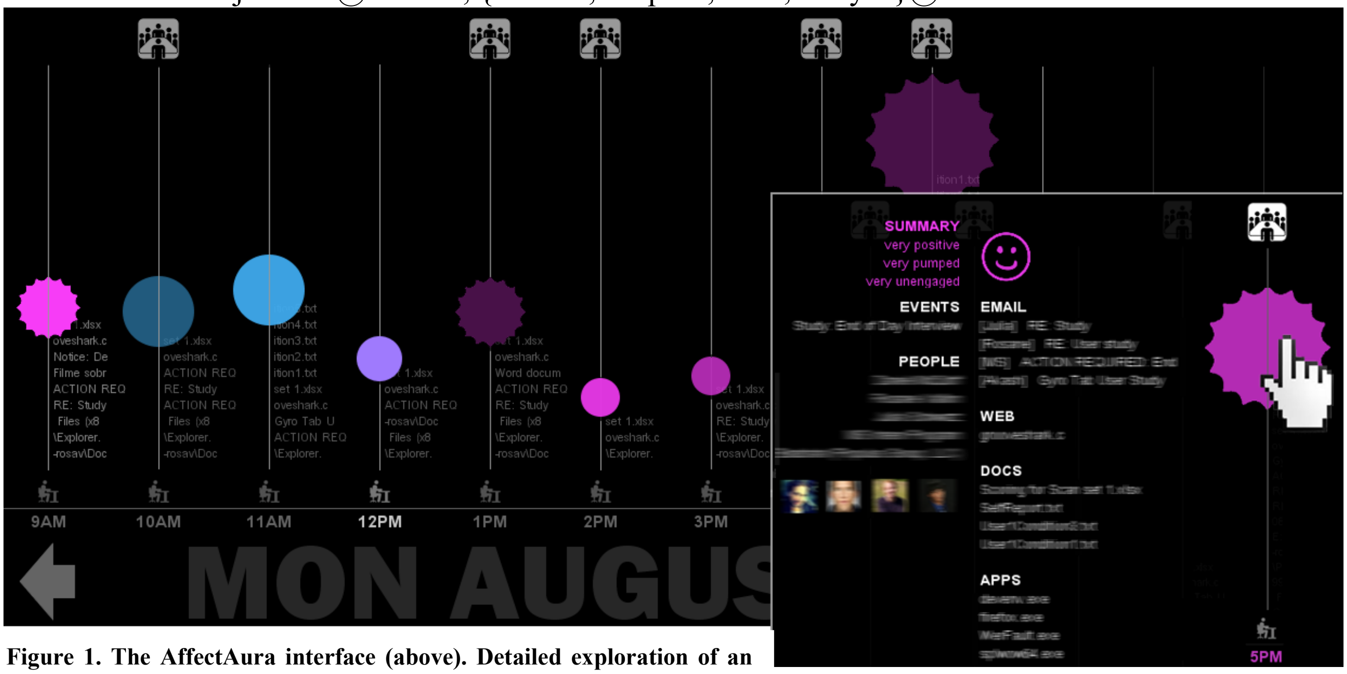

Figure 1. The AffectAura interface (above). Detailed exploration of an hour (5pm) enabled by hovering over the bubble (inset).

ABSTRACT
We present AffectAura, an emotional prosthetic that allows users to reflect on their emotional states over long periods of time. We designed a multimodal sensor set-up for continuous logging of audio, visual, physiological and contextual data, a classification scheme for predicting user affective state and an interface for user reflection. The system continuously predicts a user's valence, arousal and engagement, and correlates this with information on events, communications and data interactions. We evaluate the interface through a user study consisting of six users and over 240 hours of data, and demonstrate the utility of such a reflection tool. We show that users could reason forward and backward in time about their emotional experiences using the interface, and found this useful.

Author Keywords
Affect; Visualization; Machine Learning.

CHI'12, May 5–10, 2012, Austin, Texas, USA.

ACM Classification Keywords
H.5.2. User Interfaces: Graphical user interfaces (GUI); I.2.6. Artificial Intelligence: Learning

INTRODUCTION
How accurate are we in terms of reflecting on our everyday emotional experiences? While we may expect that we are fairly good at assessing our current emotional state, or on average how we have been feeling over the last day or so, do we really accurately assess the highs and lows over the course of a week or month or longer? And, might it be interesting to correlate the spectrum of emotions we experience with particular activities, people or places? One can imagine that a targeted assessment of one's emotional state could be valuable when considering health issues, mental health issues, eating habits, and socializing habits, to name a few. Even dramatic and stressful life changing events might be handled differently if the apparent or expected toll could be reviewed or estimated from previous experience.

With the availability of an increasing number of ubiquitous sensors it is becoming possible to more accurately document activities and make predictions about a user's future actions, behaviors and internal state. This is reflected in the growing numbers involved in the 'Quantified Self' movement [28], which is characterized by members who collect detailed long-term data about themselves, through a variety of sensors and devices, with the ultimate goal of making optimal health and behavioral choices in their daily lives.

One facet of data we can log and monitor about ourselves is how we experience the world from an emotional perspective. Our so-called *affective state* can have a significant impact on our day to day lives, not the least in the decisions we make [24]. In capturing affect, we need to consider that affect is a general term that encompasses several experiential dimensions. One of the most popular models of affect includes dimensions of valence and arousal [22]. It is widely held that valence (whether it is a positive or a negative state) is important in decision making [18] whereas arousal (the intensity or excitement level) has significant impact on memory [6]. Over long periods of time there is great potential utility in measuring these dimensions of affect, such as for capturing significant events that may affect decision making, but which may not be readily remembered. In addition to valence and arousal, the cognitive state of engagement might help users discover mental and behavioral patterns associated with episodes of high and low productivity. To explore this possibility, we tackle the challenge of tracking states of valence, arousal and engagement and provide an interactive tool for users to reflect upon. The specific contributions of this paper are:

(1) The design of a multimodal system that continuously estimates users' state valence, arousal and engagement using a variety of non-verbal and contextual cues.

(2) The design of a reflective tool, called AffectAura, that combines automatic labels for valence (negative, neutral, positive), arousal (low, high) and engagement (low, high) with contextual information to construct and present a life log tagged with affective states for the user. While prior work has looked primarily at arousal and stress [23,30], valence has received less attention because it has been harder to calculate historically due to a lack of adequate sensing technologies or the ability to track additional variables like context, which as we will see, helps with valence estimation.

(3) The evaluation of AffectAura in a multi-day user study. We demonstrate that affect estimation and presentation are key to the user experience of an emotional life log.

RELATED WORK

Over the past 15 years, there have been significant advances in the domain of affect recognition from multiple modalities [25]. However, many of these research efforts do not evaluate the results of such an affective system from the user's perspective. Kapoor and Picard [21] developed one of the first multimodal affect recognition systems, predicting student interest in a learning environment. One of the most complete multimodal systems [9] was built to predict user activities (instead of affective states) throughout their day. Other examples are: StartleCam [17], designed to detect startle responses automatically from video, and Fogarty et al's [19] system for predicting user interruptibility. None of these examples attempted to evaluate the utility of the information for the user from the perspective of reflection. G.I.R.L.S. is a technology designed to support reflection, specifically reflection on the affective impact of stories [9].

El Kaliouby et al. [14] have discussed the idea of an emotional prosthetic using wearable sensors to aid understanding and interpretation of emotion. We extend the concept of an emotional prosthetic by combining continuous affect recognition using a multimodal sensor system with contextual data drawn from events, communications and data interactions, and conduct a multi-day user study of the results.

As stated, the purpose of our evaluation was to better understand people's reactions to a retrospective memory tool that surfaces automated interpretations of their affective experiences. While this goal shares similarities with the work of Ståhl et al. [29], our approach differs in important ways. Their Affective Diary was a tool that provided semi-automated diary entry creation by incorporating abstract representations of mood and activity (cognitive and physical experiences) together with data streams from the user's phone, and users' own hand-written entries. The tool was designed for the active daily creation of diary entries, and the authors studied how incorporating affect enhanced the creation experience. Our system, in contrast, is designed as a technology probe to explore the potential reflective power that might be offered by pairing affective data with knowledge of workers' information and data interaction artifacts. We specifically designed transparency into the system so that participants could view and reflect upon the relative affective measures predicted by the system, namely; positive vs. negative valence, low vs. high arousal, and low vs. high engagement.

Reflection on one's internal mental and cognitive state has been shown to help us better understand our, and others', responses to stress [23,30]. Visual representations of users' activities have been used to great effect in providing intuitive portraits of a user and their behaviors [12]. We present a solution that allows automatic life logging and encourages reflection on emotional experiences more generally. In our design of AffectAura, our reflective interface was particularly inspired by Themail [31] and lessons learned in prior work [2,29] that demonstrate the power of visual abstraction in encouraging flexible, personal interpretation of affect-based data. Themail represents the changing nature of email correspondence between two parties over time. We use a similar abstraction but rather than representing email communications as bubbles we represent the user's predicted affective state in a somewhat similar way.

When looking through digital material for reminders of events past, triggers—rather than complete details—are frequently sufficient in order to recreate useful details from memory [5]. By combining affective labels with detailed contextual information we investigate this principle from the perspective of emotional events, and show that they can also serve as powerful memory aids.

USER SURVEY

We conducted a survey of 83 (37 females) men and women working at a large IT company around how well they felt they were able to recall their affective state and activity levels over the course of a day, a week, or a month. We also asked them whether they would appreciate a tool that helped them to reflect on their mood and activity levels over time. The purpose of this survey was to verify that users did feel that they had trouble tracking their emotional state and activity levels over time, and whether or not they thought an emotional diary might be useful.

Respondents represented a wide range of professional roles (e.g., designers, researchers, engineers, managers) and ages (18% were in their 20's, 49% in their 30's, 22% in their 40's, and 11% were 50 or older). On average, our survey participants felt fairly confident that they could track their emotional states in the near term, such as within a day, but were much less confident in their ability to do so over the course of a week or month (Figure 2). In addition, it was clear from the survey respondents that they believed there were key triggers to their emotional states. In addition to people, places and activities there were many more important triggers offered to us by participants. For instance, physical exercise and outdoor activities, playing music and sunshine or weather were most frequently cited. The majority of the survey respondents thought an affective life-logging tool might be beneficial for reflective purposes, motivating our efforts.

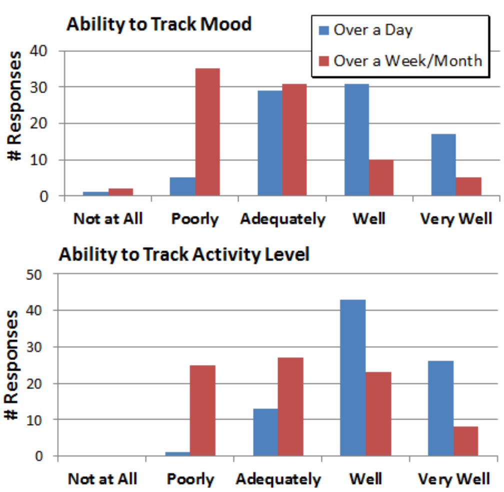

Figure 2. Survey results showing the ability with which people feel they can track their mood (top) and activity levels (bottom) over a day vs. a week/month.

LIFE LOGGING

Given our goal of developing a visualization tool for aiding user reflection on emotional states, our first task was to use Machine Learning (ML) to develop an affect recognition engine based on sensed data. Drawing from past research on appropriate end-user signals for capturing user affect, we developed a suite of desktop-based and body-worn sensors. A multimodal sensor setup using a combination of portable and workstation-based sensors was created to collect data for prospective users. We describe the individual components of the system below:

Webcam: Facial actions and expressions have been studied widely in the context of affect recognition. Facial actions such as smiles, frowns, eyebrow raises, head nods and shakes can be good indicators of valence (positive vs. negative). In order to capture such activity, a webcam was used to analyze facial actions and head motion while the participant was at their desktop. An active appearance model (AAM) [7] was used to track 100 feature points on the face, Euler angles for pitch, jaw and roll of the head and x, y and z displacement of the head within the frame. The webcam was mounted on top of the participants' computer screen. The sampling rate was 5fps.

Kinect: Posture has been found to be a good indicator of interest levels in prior work on affect recognition [21]. For example, a person highly engaged in a task can be expected to maintain still upright posture, when compared to the fidgets and slouched posture of a disengaged person. We hoped to capture such dynamics using a Kinect device. Specifically, a Kinect was used to record posture features from the participant while they were at their desk. The depth information and skeletal tracking [27] was used to determine direction of lean (left-to-right and forward-to-back) of the participant. The sampling rate was 1fps.

Microphone: Speech can be a very rich modality in terms of affect information it contains. For example, the change in prosody, structure of pauses, relative volume of the subject's speech, the exhalation of breaths can be very indicative of the person's internal state and can provide important cues both about arousal as well as valence [15]. A microphone within the Kinect sensor was used to record a raw WAV file of the audio at the workspace.

EDA Sensor: Electro-dermal activity (EDA) is widely considered one of the strongest features that can be used for detecting arousal [3]. Our sensor suite included a wearable wrist sensor [26] to record EDA and three axis acceleration of the wrist of the participant throughout each day. The sampling rate was 8Hz.

GPS: Based on the initial survey, we expected location to be an important variable associated with the affective state of the users. Consequently, a portable GPS device was carried by the participants to record latitude and longitude data. The sampling rate was 0.1Hz.

File Activity: The task that a user is working on can be fairly informative about a user's level of engagement as well as their affective states. This is also supported by the initial survey where more than 60 respondents mentioned that the projects they work on can act as a trigger to put them in a positive mood. A custom logging application was used to record file activity at the user's desktop, including web URLs visited, documents opened, applications used and emails sent and received.

Calendar Scraper: Finally, the people we interact with are one of the most important components of affective states we experience. Our system captures people interactions using calendar information extracted using custom software; in particular, we extract times and attendees of meetings scheduled in users' calendars. An example of the desktop set-up for one of the participants is shown in Figure 3.

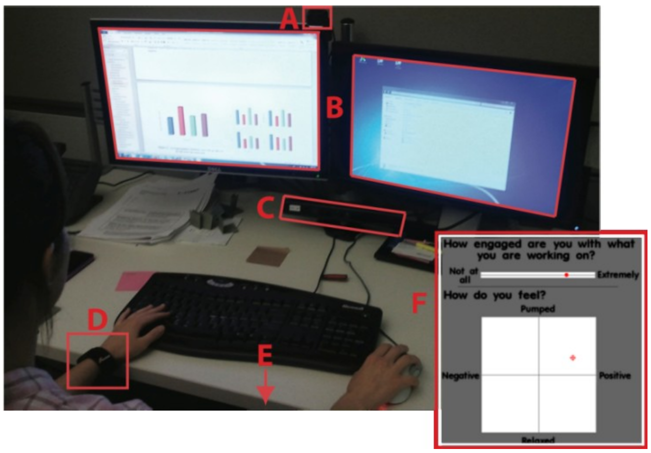

Figure 3. Sensor configuration: A) webcam, B) file activity and calendar monitor, C) Kinect, D) EDA sensor, and E) GPS. F) The self-report interface.

Using Machine Learning for Recognizing Affect

To provide a system for surfacing affective signals to the user, we required a way to take the multiple streams of data we described above and output a prediction across the three dimensions of valence (negative, neutral, positive), arousal (low, high) and engagement (low, high); that is, we needed to develop an affect prediction engine. To develop the predictive models necessary for the affect prediction engine, we gathered data from five participants and used well-known machine learning methods to build models for mapping raw multi-sensor streams to discrete values for valence, arousal and engagement.

Specifically, the sensor streams were each normalized and features were extracted as detailed in Table 1. The data was divided up into 10 minute segments. To train the model to predict the affect labels, we collected training data and self-report labels from five participants over two days each. To obtain affect labels for training, the participants reported their valence, arousal and engagement levels at regular intervals during the day. In this phase, labels consisted of numerical values between -1 and +1 for the three dimensions of valence, arousal, and engagement, collected using the custom tool shown in Figure 3, similar in design to that used by Dietz and Lang [11]. The labels were normalized along each dimension and thresholded to form discrete classes (e.g., positive vs. neutral vs. negative, low vs. high). We collected 110 training examples, each of which corresponds to features extracted from a 10 minute segment.

We chose to use a simple nearest-neighbor classifier in order to make the predictions. In particular, any new incoming 10 minute “test” segment can be classified for valence, arousal and engagement by finding the closest annotated 10 minute segment already present in the training dataset; the test segment thus inherits the affective annotation of the closest match found. For this scheme to work well requires a method for computing distances between the features of the 10 minute episode, which is optimally accomplished using a distance-metric learning technique. In particular, we use Neighborhood Component Analysis (NCA) [16]. Intuitively, the goal behind NCA is to weight different features such that the examples belonging to the same class have a minimal distance, while the ones belonging to different classes are as far apart as they can be. In our system, we learn a separate distance-metric for each of the three different labels (arousal, valence and engagement).

How Good is the Affect Recognition Engine?

In order to make sure that our recognition engine would work, we performed experiments to evaluate its classification performance. Our approach to training and testing the classifier is consistent with accepted practice. In particular, we employed a leave-one-out validation scheme. The idea behind leave-one-out analysis is to select a data point from the training set and then attempt to predict its label using the rest of the examples. This process is then repeated for all the examples present in the training data and the aggregate results therefore reflect the predictive capability of the ML system. Data used for validating the classifier was different from that collected for the user study described later.

We found that such a leave-one-out analysis resulted in an overall accuracy of 68% across all three states and indicates that our scheme does have the ability to predict the affective states. The system was able to predict the state of engagement most accurately (71%). Note that, these affect recognition results are on par with existing systems that operate on natural data, recorded mostly in uncontrolled environments [21] (as opposed to enacted ones recorded in a lab environment). Also note that for our application purposes, we do not require an affect recognition engine that is 100% accurate. AffectAura simply uses the output of the predictor for purposes of guiding the visualization and can thus benefit from a system that has any predictive information.

Finally, it is interesting to analyze our ML model and see what stream of data matters in terms of predicting affect. In particular, we can look at the weights of the input feature space that NCA finds most significant for each class label. High weights correspond to more important features, thus they can provide some insight into the importance of different modalities. Figure 4 shows the total weight each modality was assigned in this mapping after performing NCA. We can observe that EDA sensor (EDA + Acceler.) and posture (Kinect) information are most useful for predicting arousal and engagement respectively. In addition, of the three affect dimensions, head displacements, rotations and facial activity (AAM features) contributed the most weight to the valence prediction. These observations are in line with our intuitive beliefs about the importance of different channels and reinforce the benefit of using a range of sensors to capture diverse information for affect classification.

AFFECTAURA'S INTERFACE

AffectAura is a visualization of the user's estimated affective states over time. It incorporates two components 1) the affect prediction engine for labeling users states and 2) a timeline interface, surrounded by other context, for the user to reflect on these data. Our intention was to create a compelling experience that might engage the user to reflect on their emotional state throughout the day. We were inspired by many timeline visualizations [1] of activities and data.

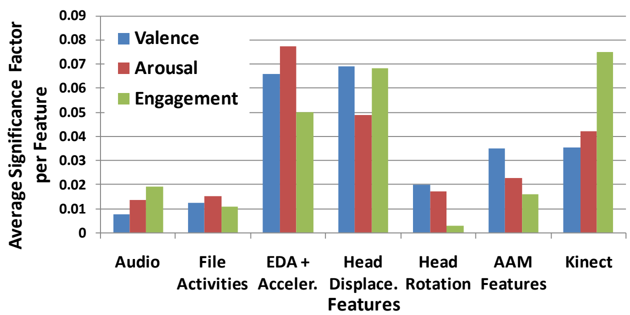

Figure 4. The factors each feature contributed to the final NCA mapping for each of the three classes. EDA sensor features were given the most weight in total.

AffectAura consists of an 'affect timeline' capturing the ebb and flow of affect represented by a series of bubbles. The user can drill down by hovering over a bubble to reveal a detailed breakdown of their activities and interactions associated with that time. We chose to display daily and hourly granularities of data for the user, to simplify the visualization. Each "page" shows one day divided into hour intervals (columns). An example of AffectAura is shown in Figure 1.

Valence of the affective state is represented by the **color** of the bubble. We used a "hot/cold" metaphor as our starting point, as emotions are often characterized as hot or cold, but ended up with pink, purple and blue representing positive, neutral and negative affect, respectively.

Arousal of the user is represented by the **shape** of the bubble, a circle or a "burst" that were chosen to represent calm and "pumped up" respectively. This type of mapping has been shown to be effective in the SAM (Self-Assessment Manikin) representation [4].

Engagement of the user is represented by **opacity**, with higher opacity corresponding to greater engagement.

Figure 5. Annotated representation of an hour.

We iterated through many visualizations of timeline events, such as "streams", bars and various other shapes but the bubbles provided multiple parameters (size, shape, color and opacity) that could represent different dimensions of the affective states and were most aesthetically pleasing. The representations of engagement and arousal changed significantly from size and motion to opacity and shape for clarity and aesthetic reasons during the iterations.

For each hour, the user's desktop activity (file and application activations) is represented by the size and height of each bubble (radius and height were linear functions of the number of activities), as shown in Figure 5. A subset of the activities themselves are listed as a ghosted column of text to lend a sense of the amount of file and application interaction that occurred. This text list representation was inspired by the Themail visualization of email keywords [31].

If the user had a meeting scheduled in their digital calendar during a particular hour, a meeting icon was displayed at the top of the column. Similarly, the location of the user for

Sensor	Stream	Normalization	Features (calculated for every signal window)
Webcam	AAM Features	Used Procrustes analysis [8] to remove 2D variations due to rotation, displacement and scaling.	Mean smile intensity, Mean eyebrow activity, Mean mouth activity
	Rotations	Remove mean from axes	Pitch, yaw and roll (standard deviation)
	Displacements	Remove mean from axes	X, Y, Z displacement*
Mic	WAV audio file	N/A	Length of conversation, no of turns in conversation, zero-crossing rate*, normalized frame energy*, gradient index*, local kurtosis*, spectral centroid* [20]
Kinect	Depth frame	Remove mean	Gradient of lean (front-to-back)
	Skeletal tracker	Remove mean	Gradient of lean (side-to-side)
EDA Sensor	EDA	Remove mean and low pass filter (cut-off 0.4Hz)	Mean, gradient, different between first value and max, position of max, difference between value and min, position of min, zero crossings, number of peaks.
	3-axis accelerometer	Remove mean and unit std for each axis	Standard deviation.
File Activity	File instances	N/A	No of activities, no of unique activities

Table 1. Table showing the sensors used, data streams recorded, the method of normalizing each data stream and the features calculated from the normalized streams. *Mean and standard deviation was calculated for these values. Calendar events and attendees and GPS latitudes and longitudes were also collected for context but are not used for classification.

a particular hour, based on GPS and file activity, was represented as an icon at the bottom of the column; three icons were used, representing 'home', 'work' and 'elsewhere'. These icons were included to aid recall of events. The user could navigate to different days by clicking on the left and right arrows at the bottom of the screen.

To allow for exploration of the events of each hour, the user could hover over any bubble to reveal details of their activity logged during that hour (Figure 1, inset). The artifacts presented on hover included calendar events, people (names and when available, images), email titles (with sender name), website URLs, document names, and applications used. All artifact data presented was gathered via our custom logging application and calendar scraper utility. Additionally, a text summary of the predicted affective state was displayed at the top of the column for quick perusal. AffectAura was written in Processing, (www.processing.org), an open source environment suitable for rapid prototyping.

USABILITY STUDY

We conducted a study using AffectAura as a probe to understand the reflective opportunities afforded by visualizations that pair event artifacts with predictions of user affect. Inspired by our survey, which suggested that memories of events and emotions drop off rapidly after a day, we targeted the work-week as an interesting multi-day period of time over which to investigate memory and reflection.

Method

Each participant recorded data using the sensor set-up described previously for four consecutive days, Monday to Thursday of a working week. At the end of each day, the participants were asked to complete a short survey reporting the single most positive and negative emotional experience that day (if any) along with the magnitude of the valence of the experience on a 7-point scale (1=Negative, 7=Positive), as well as the people (maximum 5) with whom they had significant interactions that day, again assigning an emotional magnitude to the interaction. On the subsequent Friday, participants completed another short survey about the emotional events and significant interactions they had with people that week, and took part in a semi-structured interview in which they first explored AffectAura using a think-aloud protocol and then answered questions related to aspects of AffectAura that did (and did not) help them recall events of the week and associated affective responses. The evaluation concluded with a survey that gathered participants' subjective reactions to the usability of the AffectAura visualization. Total evaluation time was 60 minutes.

Participants

Six participants (3 females) took part in the evaluation of AffectAura. Their ages ranged from 26 to 33 (Mean: 28.5). Four were graduate students and two were professional researchers. Our only screening requirement was that participants regularly spent at least four hours a day at their desk. On average, participants recorded over 10 hours of sensed data per day, totaling over 240 hours of logged data. Participants were compensated with a $250 gift card.

Results

Emotional memory

To understand how the passage of time impacted participants' memories of emotions and events experienced during the week, we compared survey responses that participants gave on interview day (Fri) to the responses they gave on each day of data collection (Mon-Thurs). We compared participants' memories of the people they interacted with, the most positive and most negative event that happened each day, and the emotional charge associated with each.

Considering first the participants' memories of interactions with people, participants generally recalled more people that they interacted with two days prior to the interview (Wednesday, μ=3.2; Thursday, μ=3) than from the beginning of the week (Monday, μ=2; Tuesday μ=2). Overall, the vast majority of social interactions were considered relatively positive (79%) vs. relatively negative (8%) on the day of the event, and participants were quite good at recalling the relative emotional charge of the interaction; out of the 62 people interactions that participants correctly recalled, the emotional scale value (1-7) of the interaction was correctly recalled 63% of the time. In the 23 cases where participants did not recall the precise emotional value of the interaction, they were off 1 point in all but one case, thus correctly recalling the general tone of the interaction. Participants were twice as likely (15 instances) to misremember the exact tone of interaction slightly more positively than slightly more negatively (8 instances). In the 1 case that a participant did not correctly recall the negative tone of interaction, the participant had had several positive interactions with that person during the week, which we hypothesize overshadowed the one negative interaction. This interpretation agrees with our interview with one participant, who claimed that she generally associates a certain tone of interaction with a person. However, this may suggest that exceptions to the norm are worth noting.

Participants recalled a similar percentage of positive and negative events from the prior four days (55% of the 31 total events) as they did their significant people interactions (56% of 111 people). As with people, on average, participants recalled more events from the previous day (Thursday, μ=1.5 events per participant) than from the beginning of the week (Monday, μ=0.8 events per participant), with a curious dip in recollection on Tuesday (μ= 0.2). Of the events recalled later in the week (Wednesday and Thursday), 81% of those that had a negative tone (<4) were recalled, versus only 40% of the ones with a positive tone (>4). The opposite trend was seen for Monday events, with participants recalling 67% of the positive events versus only 25% of the negative events. This supports evidence that memory of unpleasant events fades faster than that of pleasant events [32], and probably reflects the effects of primacy/recency on long-term memory [13].

We also asked participants to recall the overall emotional tone of each day of the week, and to rate their confidence in their recollection on a 7 point scale (1="not very confident", 7="very confident"). Participants reported that their days were positive (a score of greater than 4) 66% of the time, negative 25% of the time, and neutral otherwise. While participants were more confident in their recollections of the previous 24 hours (Thursday μ=5.2) than from early in the week (Monday μ=4.3), participants' confidence levels did not lead to better recollection. We saw particularly poor performance for recalling details about Tuesday, with 5/6 participants misremembering the exact tone of the day and 3 misremembering the general tone of the day (e.g., recalling a generally positive day as negative and vice versa). Since participants did relatively well in recalling the general positive or negative tone of prior days (83% correct), perhaps the most useful observation from this data is that people can suffer lapses in memory about the emotional tone of a day in as little as 24 hours.

Using AffectAura

Overall, our study participants were successful in reconstructing a "story" of their workday activities by exploring and reasoning over the data artifacts presented in AffectAura (e.g., events, people, emails, web sites, documents and applications). Gaps in the data we collected sometimes made it difficult for participants to determine what they were working on in a given hour; for example, participants could not always discern what document or application they were working with. Even so, participants could often reason forward and backward in time based on information presented earlier or later in the day, which allowed them to piece together the major activities of the day and to reason about whether the affective signals resonated with them. We consider this a novel and impactful contribution to the life logging and 'Quantified Self' movement efforts.

Recalling Forgotten Events with AffectAura

During exploration of the interface, all six recalled at least one event, generally involving a person that they had forgotten about at the time of the interview, and agreed that a system that helped them remember the event and tone of the event at a future date would be useful. However, comments by four participants suggested that not all interactions with people are created equal, and that events worth remembering are those that are out of the ordinary:

> *"If there was a visitor that day and I had a very nice interaction with that visitor, and [months] later, I want to recall what are the important highlights of my week, yes, that would be helpful to recall. But these people I interact with every day, I already know that I like him, or that I like interacting with that person, so it wouldn't really help me that much" –P1.*

In addition, a quote by P5 suggests that what qualifies as "out of the ordinary" may be based on many factors, including the frequency of interaction with a known person, meeting a new person, using a different medium to interact with the person, or even an unusual topic of discussion:

> *"For a person whom I'm interacting very regularly with like my manager or my intern on the project, there would be so many entries I probably wouldn't be so interested. But if it's an outlier, like a visitor, or somebody I met, or an email from a friend which is not a usual email, I would be interested in remembering those […] The reason it [an IM chat with his father] stands out in memory is because I normally don't talk to my parents a lot on the chat [it] was not the regular kind of talk, so that's why I would be happy if I could remember that." – P5*

We expect that existing life logging tools and systems could readily be instrumented to detect infrequent or new interactions with people, but that the addition of affective sensing technology might be particularly useful for tagging interactions with people that trigger an atypical emotional response. The potential utility of tagging interactions based on emotion is supported by the comments of several participants who provided examples of events that would be useful to remember because of their positive or negative tone, even though it was someone they typically interacted with.

How Participants Used Affective Cues in AffectAura

Two questions we were particularly interested in understanding as we explored AffectAura with our participants was 1) whether participants used the affective information to recall events; and 2) whether the affective information resonated with their own memory of a day.

Affective Cues Aren't Memory Triggers. Through our interviews, we found little evidence to suggest that affective cues alone helped people recall events from earlier in the week. Instead, participants tended to reason about the affective information cues once they had reconstructed what they were doing that hour or day based on other artifacts. All six participants found multiple hours across the week where the activity that the interface suggested that they were doing during an hour (e.g., from events, people, emails, etc) was accurately reflected by the affective information presented, e.g., P5: "*At 11 I was in the talk and it shows me as engaged, calm and neutral, and that was definitely true*". We also saw examples that the visualization accurately reflected the overall mood of the day, again, once the participant had reconstructed the day from specific data artifacts:

> *"I'm seeing the emails I sent out and at this point people are replying to my emails, essentially pointing out things are broken and I'm realizing throughout the day that the problem is bigger than I think it is. But it could have been Wednesday when I really realized...[looking at the emails] no, it was this day....Now that I scan through the emails and remember what the day was, and *now* I look at the interface, it makes total sense that this was what was collected. Mainly because it looks like I was excited at the beginning of the day,*

> *things are going great, and there was a sudden drop off and all of a sudden I'm very negative, which is exactly what happened on Tuesday." –P4*

P4 provided another example for Thursday, where the highs and lows related to a talk he gave were accurately reflected at the day view. P2 had a similar experience: once he had reconstructed the events of Monday he commented "*I like this curve [the varying height of the bubbles indicating "activity"] and that probably feels about like my day*". Whereas the patterns over a specific day resonated strongly with P2 and P4, other participants found that the interface provided some insights about patterns across days, for example that P3 is more pumped in the morning and more positive when interacting with people and that P1 was more positive at lunchtime because of her interactions with friends.

Participants did discover examples where they could not find direct evidence for the affective information that the system was presenting. Interestingly, this often led participants to rationalize the information based on what they remembered about the day that would not have been captured by our system, such as having had morning coffee could have made P1 pumped or having ridden his bike to work could have made P2 calm. The fact that such explanations are only hypotheses suggests there may be some danger of creating false memories. This is especially true given that all participants provided examples for which the affective information contradicted their memory of an hour. A particularly introspective participant commented "*I don't know how much of this [his interpretation] is coming from the data itself or what I'm reading into it—apophenia.*" (P2).

Rationalizing 3 Dimensions of Affect Simultaneously is Hard. Despite the numerous examples that participants provided where they agreed with the positive/negative, pumped up/relaxed, or engaged/unengaged affect reflected in the visualization on a given hour, it was relatively rare that all dimensions of affect made sense simultaneously. While we intentionally surfaced the system's predictions for all three affective aspects to understand what information resonated with participants and what did not, the act of rationalizing all three clearly increased the cognitive demands of using the interface and a simpler design should be explored.

Affective Cues that Aligned with Memory Reinforced the Experience. While AffectAura's affective cues did not always aligned with a participant's memory of an event, participants typically made a simple verbal note of the disconnect and continued their exploration. By contrast, when AffectAura' affective cues did correspond with the participant's memory of the event, participants often declared the fact emphatically (e.g. P2: "*it did exactly what I recall*", P4: "*that's exactly what happened…I'm shocked you were able to collect this data so well*"; P5: "*that was definitely true*"), and took a moment to explain why the prediction was correct. Participants *enjoyed* finding examples that accurately reflected their memories. While this is an early finding, it offers intriguing evidence that capturing and reflecting on affect can create engaging retrospection experiences.

To summarize, the visualization of affect across an entire day often made sense to participants in retrospect, but was not sufficient to reconstruct the events of the day. Even so, participants appreciated the times that the system's affective information matched their memories, which seemed to contribute to a more engaging reflective experience. This helped them discover and reflect on productivity-based and emotion-based patterns in their day.

How Participants Used Data Cues

We were curious to understand whether certain data types were more (or less) useful in reconstructing events of the day, and what important types of data were missing from the interface. Drawing from both our interviews and a post-interview questionnaire that asked participants to rate the utility of each data type in helping them reconstruct their day (1="not very useful" to 7="very useful"), we found no consensus among participants; every data type except for URLs was ranked highest by at least one participant.

URLs, Applications, and Documents Didn't Work. URLS received the lowest median rating (3) across data types, perhaps best explained by P2 who found the URLs "*meaningful but noisy*". Participants also did not consider application information to be very useful (median=3.5) as it was represented in the interface. While documents were one of the highest ranked data type in our survey (median=5), participants never mentioned a document when reconstructing the week's events using the interface; we suspect the discrepancy arises from the fact that AffectAura did not display the complete list of documents for space saving reasons, but that participants *perceived* that document names accurately reflect what they were working on.

Email, Events and People Did Work. All the data types that directly or indirectly referred to people—email, (calendar) events, and people—featured prominently in helping participants reconstruct the events of the week, and all were rated similarly in terms of their utility (median=4). All agreed that email would be more useful if it were prioritized by sender, recipients or time. Even so, in 6 instances, email played a critical role in helping participants recall events from the week. Calendar events were particularly useful in helping participants tell a "story" about a particular day. Calendar events were also strong anchor points in helping participants recall their mood, engagement level and energy level; without exception, when a participant reflected upon a calendar event, they were able to judge whether or not the affective information corresponded to their mood

Given the role that people play in both email and calendar events, it is not surprising that people were by far the most often mentioned data type as participants explored the interface. The power that a person's name can hold in helping users recall past events is described by well by P1:

> *"I recall it now, but three months from now I wouldn't be able to recall [that I interacted with] that person .... it brings back the sequence—there is a story behind that interaction. So seeing that name brings back all the context of that day." – P1*

Participants extracted a surprising amount of value from behavioral information about when they were and were not at work. For four participants, seeing the work pattern (e.g., arrived late, left early) not only triggered a previously forgotten memory of an event, but was sufficient to recollect several other events that happened that day. This provides further evidence that exposing and prioritizing "out of the ordinary" events may be valuable in a retrospective system.

How Participants Used the Visualization Features

Drawing from participant feedback as well as our observations as participants explored the AffectAura interface, we developed insights into the design considerations for future systems that support retrospective memory.

Data Granularity. One tension was the interface providing too much data to make sense of at a glance, but too little data to reconstruct the activities of every hour, or even some of the major events of the day. When participants were successful at reconstructing a "story" of the day, it was mostly because they discovered one or two cues while sifting through the details of the day that they could then use to reason forward and backward in time. Participants cited important streams of data that were not captured by AffectAura, such as face to face interactions, chat/IM/skype history, web-based email, phone calls and text messages.

While future systems can certainly strive to include ever more data streams to offer a more complete picture of people's days, participants' comments suggest that it is equally important to reduce the complexity of the interface. For example, web and email lists included unimportant items that masked the meaningful ones. Consider P3, who on the one hand appreciated AffectAura ("*I think it was great because I usually don't recall details. But looking through here I remembered things. Even if it wasn't data in the UI it helped me remember facts of the week*"), was also overwhelmed by the volume of data presented ("*daily, this is too much*"). This sentiment was echoed by P6, who found that "*going through the day was tedious.*"

Such user comments highlight the need for judicious presentation of the available data. Most participants suggested that they did not want to *give up* the details, but that higher levels of abstraction coupled with drill down would be appropriate. Furthermore, two participants explained why even more detail might be required at the lowest level, since an hour was too coarse a timescale to reason about certain events (e.g., meaningful events can get swamped by less meaningful events), and that the boundaries of an hour were too rigid (e.g., meetings that start halfway through an hour have little to do with the deskwork that was done in the first half of the hour.) It remains an open question as to whether adding more detail would support users in discovering the precise source of an emotional experience. This would require robust, fine-grained prediction coupled with "complete" data streams, neither of which is possible today.

Cue Mappings. The visual cues for arousal and valence made sense to most participants. Size/height related to desktop activity did not resonate with most participants as they preferred out of the ordinary events, which were not always correlated with desktop activity, to be highlighted.

How Participants Anticipated Using Such a System

The interviews highlighted a number of potential uses for a system like AffectAura. Firstly, as a memory aid to record events automatically and help the user recall events upon reflection. In particular, the serendipitous recall of interesting or special events was highlighted by two users. The ability to edit entries was raised as an important feature.

The use of the system as a search tool, not purely for reflection, was raised by two users. This was tied to the ability to better understand patterns within the affective and contextual data on which four of the users commented. Examples given were of the relationship between physical activity and productivity to one's affective state and more generally patterns of stress over time. Most participants said they would be interested in this type of data weekly or monthly, reflecting their desire to use this information retrospectively to aid memory, which drops off quickly after a day or two.

The study highlighted challenges for future iterations of such a design. One user indicated that over longer periods of time it would be important for the system to discern between the highs and lows of a day (which may be relatively insignificant) and the highs and lows of a month or year.

> *"The highs and lows of this day were not even high enough and low enough to be memorable. Although this was a bad moment, and although I may remember it ... that's a blip on the radar." – P2*

FUTURE WORK

In this work most of the sensors and contextual information was limited to data that could be collected at the participant's desktop. Sensors could be utilized in order to capture significant emotional experiences and memory triggers may occur in other contexts. In this work we did not explore additional classification techniques for predicting the users' states. More comprehensive training examples and person-specific training data could improve the classification results. Several users commented that being able to view the data at more granularities may be desirable. A system that detected abnormal patterns of behavior could also be useful. Finally, users could assist the system with prioritization learning.

CONCLUSION

We have presented AffectAura, an emotional prosthetic that allows users to reflect on their emotional states over time. We designed a multimodal sensor set-up for continuous logging of audio, visual, physiological and contextual data, and a classification scheme for predicting affective dimensions, including valence, arousal and engagement. We presented and evaluated an interface for user reflection using the system. We demonstrated that users were able to leverage cues from AffectAura to construct stories about their days, even after they had forgotten these particular incidents or their related emotional tones. Five of the six participants told us that affect would be useful for reflecting on events collected in the user interface. We believe this is the first automatically collected, longitudinally evaluated emotion memory system. We have delineated many benefits to emotion and activity tracking and the particular visualization we chose to expose the information to users.

REFERENCES

1. Aigner, W., Miksch, S., Schumann, H. and Tominski,C. Visualization of Time-Oriented Data. Springer, 2011.
2. Boehner, K., DePaula, R., Dourish, P. and Sengers, P. Affect: from information to interaction. *In Proc. CC '05*, ACM (2005), 59-68.
3. Boucsein, W. *Electrodermal Activity*. Springer-Verlag, New York, 1992.
4. Bradley, M.M. and Lang, P.J. Measuring emotion: the self-assessment manikin and the semantic differential. J *Behav Ther Exp Psy*, 25, 1 (1994), 49-59.
5. Brush, AJ, Meyers, B.R., Tan, D.S. and Czerwinski, M. Understanding memory triggers for task tracking. *In Proc. CHI '07*, ACM (2007), 947-950.
6. Cahill, L. and McGaugh, J.L. A novel demonstration of enhanced memory associated with emotional arousal. *Conscious Cogn*, 4 (1995), 410-421.
7. Cootes, T.F., Edwards, G.J. and Taylor, C.J. Active appearance models. *IEEE Pattern Anal* (2001), 681-685.
8. Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J. et al. Active shape models-their training and application. *Comput Viz Image Und* (1995), 38-59.
9. Daily, S. B. and Picard, R. W. Girls Involved in Real Life Sharing: Utilizing Technology to Support the Emotional Development of Teenaged Girls*, J School Counseling,* 5,1(2007), 20.
10. De la Torre, F. and Agell, C. Multimodal Diaries. *In Proc. ICME '07* (2007), IEEE, 839-842.
11. Dietz, R. and Lang, A. Affective agents: Effects of agent affect on arousal, attention, liking and learning. *In Proc. COGNITIVE '11*, (1999).
12. Donath, J., Dragulescu, A., Zinman, A., and Viegas, F. Data portraits. *In Art Gallery ACM SIGGRAPH* (2010).
13. Ebbinghaus, H. (1885). Memory: A Contribution to Experimental Psychology. *Tran.Ruger & Bussenius(1913).*
14. El Kaliouby, R., Teeters, A. and Picard, R.W. Invited Talk: An Exploratory Social-Emotional Prosthetic for Autism Spectrum Disorders. *In BSN '06*, (2006), IEEE.
15. Fernandez, R. A computational model for the automatic recognition of affect in speech. Massachusetts Institute of Technology, 2004.
16. Goldberger, J. and Roweis, S. and Hinton, G. and Salakhutdinov, R. Neighbourhood Components Analysis. *In Proc. NIPS '04*, MIT (2004).
17. Healy, J. and Picard, R. W. StartleCam: A cybernetic Wearable Camera, *In Proc. 2nd International Symposium on Wearable Computers*, 1998.
18. Higgins, E.T. Beyond pleasure and pain. *Am Psychol* (1997), 1280.
19. Fogarty, J., Ko, A.J., Aung, H.H., Golden, E., Tang, K.P. and Hudson, S.E. Examining task engagement in sensor-based statistical models of human interruptibility. *Proc. CHI '05*, ACM (2005), 331-340.
20. Jax, P. and Vary, P. On artificial bandwidth extension of telephone speech. *Signal Processing* (2003), 1707-1719.
21. Kapoor, A. and Picard, R.W. Multimodal affect recognition in learning environments. *In Proc. MULTIMEDIA '05* (2005), ACM, 677-682.
22. Larsen, R.J. and Diener, E. Promises and problems with the circumplex model of emotion. *Pers Soc Psychol Rev* (1992), 25-59.
23. Liu, K.K.L. A personal, mobile system for understanding stress and interruptions. MIT, 2004.
24. Loewenstein, G. and Lerner, J.S. *The role of affect in decision making*. In Handbook of affective science. Oxford University Press, Oxford, 2003.
25. Nicolaou, M.A., Gunes, H. and Pantic, M. Continuous Prediction of Spontaneous Affect from Multiple Cues and Modalities in Valence-Arousal Space. *IEEE T Affect Comput* (2011).
26. Poh, M.Z., Swenson, N.C. and Picard, R.W. A wearable sensor for unobtrusive, long-term assessment of electrodermal activity. IEEE T Bio-med Eng (2010).
27. Shotton, J., Fitzgibbon, A., et al. Real-time human pose recognition in parts from single depth images. In Proc. CVPR '11, IEEE (2011), 1297-1304.
28. Singer, E. The Measured Life. Technology Review (July/August 2011).
29. Ståhl, A., Höök, K., Svensson, M., Taylora, A. and Combetto, M. Experiencing the affective diary. Pers *Ubiquit Comput* (2009), 365-378.
30. Suls, J., Green, P. and Hillis, S. Emotional reactivity to everyday problems, affective inertia, and neuroticism. *Pers Soc Psychol B* (1998), 127.
31. Viégas, F., Golder, S. and Donath J. Visualizing email content: portraying relationships from conversational histories. *In Proc. CHI '06*, ACM (2006), 979-988.
32. Walker, W.R. and Skowronski, J.J. and Thompson, C.P. Life is pleasant-and memory helps to keep it that way! *Rev Gen Psychol* (2003), 203.