Memento: An Emotion-driven Lifelogging System with Wearables

SHIQI JIANG, Nanyang Technological University
ZHENJIANG LI, City University of Hong Kong
PENGFEI ZHOU and MO LI, Nanyang Technological University

Due to the increasing popularity of mobile devices, the usage of lifelogging has dramatically expanded. People collect their daily memorial moments and share with friends on the social network, which is an emerging lifestyle. We see great potential of lifelogging applications along with rapid recent growth of the wearables market, where more sensors are introduced to wearables, i.e., electroencephalogram (EEG) sensors, that can further sense the user's mental activities, e.g., emotions. In this article, we present the design and implementation of Memento, an emotion-driven lifelogging system on wearables. Memento integrates EEG sensors with smart glasses. Since memorable moments usually coincides with the user's emotional changes, Memento leverages the knowledge from the brain-computer-interface domain to analyze the EEG signals to infer emotions and automatically launch lifelogging based on that. Towards building Memento on Commercial off-the-shelf wearable devices, we study EEG signals in mobility cases and propose a multiple sensor fusion based approach to estimate signal quality. We present a customized two-phase emotion recognition architecture, considering both the affordability and efficiency of wearable-class devices. We also discuss the optimization framework to automatically choose and configure the suitable lifelogging method (video, audio, or image) by analyzing the environment and system context. Finally, our experimental evaluation shows that Memento is responsive, efficient, and user-friendly on wearables.

CCS Concepts: • **Human-centered computing** → **Ubiquitous and mobile computing systems and tools**; • **Hardware** → *Sensor applications and deployments*;

Additional Key Words and Phrases: Wearable, lifelogging, EEG, emotion recognition

ACM Reference format:
Shiqi Jiang, Zhenjiang Li, Pengfei Zhou, and Mo Li. 2019. Memento: An Emotion-driven Lifelogging System with Wearables. *ACM Trans. Sen. Netw.* 15, 1, Article 8 (January 2019), 23 pages.
https://doi.org/10.1145/3281630

1 INTRODUCTION

Lifelogging is a technique to digitize human daily lives, which was widely adopted in the therapy for a series of neurodegenerative diseases using dedicated devices [21, 22, 34]. Later, due to the

This work is support by Singapore MOE Tier 2 grant MOE2016-T2-2-023, Tier 1 grant 2017-T1-002-047, NTU CoE grant M4081879. This work is also supported by an ECS grant from Research Grants Council of Hong Kong (Project No. CityU 21203516), and a GRF grant from Research Grants Council of Hong Kong (Project No. CityU 11217817).
Authors' addresses: S. Jiang, P. Zhou, and M. Li (corresponding author), School of Computer Science and Engineering, Nanyang Technological University, 50 Nanyang Avenue, 639798 Singapore; emails: {sjiang004, pfzhou, limo}@ntu.edu.sg; Z. Li (corresponding author), Department of Computer Science, City University of Hong Kong, 83 Tat Chee Avenue, Kowloon, Hong Kong, China; email: zhenjiang.li@cityu.edu.hk.

1550-4859/2019/01-ART8 $15.00
https://doi.org/10.1145/3281630

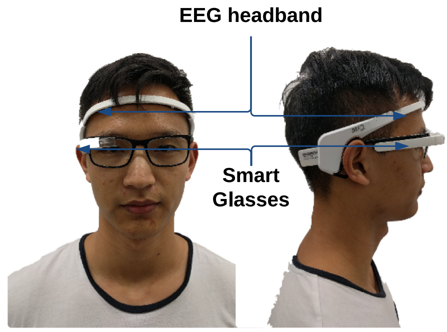


Fig. 1. Illustrataion of an emotion-driven lifelogging system.

increasing number of smartphones, the use of lifelogging has been dramatically expanded. It has further became an emerging lifestyle for people to collect their memorial moments and share with friends. Lifelogging applications or services [3, 4] on smartphones are able to log users' lives in various forms, e.g., texts, images, audio clips, videos, and so on. Recently, the wearables market has been rapidly growing in terms of both technology advances and penetration. Wearable devices, especially smart glasses, are equipped with the first-person camera, microphone, and rich on-board sensors. They are always carried by the users and exposed to the ambient environment, which can thus serve as a more suitable platform for the lifelogging service. Although prior attempts have been made for designing lifelogging systems, the proposed solutions have two major limitations:

(1) *Dependence on human intervention.* Most existing lifelogging designs either conduct the lifelogging continuously over the time or reply on manual operations [3, 4]. The limitations are twofold. First, it is energy inefficient and wastes the device's storage, because usually not everything needs to be logged. Second, due to the lag off manual operations, many brief but valuable moments could be easily missed. Though various user interaction methods are introduced on wearables, like gesture, voice, and even wink controls, the intervention overhead causes non-negligible burdens and may impair users' willingness to conduct lifelogging. Thus, an automatic lifelogging service without requiring user intervention is expected.
(2) *The affordability of wearables.* Many prior lifelogging service designs rely on dedicated devices [2, 5], which, however, are not appropriate for most Commercial off-the-shelf (COTS) mobile or wearables devices. On one hand, due to the limited size, the energy resources of devices are bounded, which cannot afford extensive and continuous sensing. On the other hand, even intelligent lifelogging could lead to massive irrelevant lifelogs, which can rapidly occupy device's memory. Therefore device's affordability should be carefully considered.

To overcome above limitations, the key motivation of this article is to automatically trigger the lifelogging service with an efficient and lightweight design. To this end, we propose Memento, a lifelogging system on COST wearable devices to automatically log memorial moments by understanding the user's mental state changes—lifelogging is triggered automatically by the user's emotions. As shown in Figure 1, Memento integrates electroencephalography (EEG) [7] electrodes with smart glasses, where EEG measures brain waves. By leveraging the Brain-Computer-Interface (BCI) domain knowledge, Memento derives emotions from EEG signals and launches the lifelogging process automatically based on that. The development of such a service, however, entails several crucial challenges.

First, the signals collected from wearable EEG sensors are not always reliable—EEG signals are usually mixed with various external interference signals from electric appliances and interfering signals from the muscle activities such as blinks, the jaw, or the heart beat. However, the electrodes of wearable EEG are not fixed on the scalp (see Figure 1 as a reference). The movements of the user could thus cause a drift of the electrodes, which will in turn lead to a unpredicted change of the harvested EEG signals. It is non-trivial to distinguish that the observed EEG signal changes are due to such drift-caused "noises" or the user's actual emotions. In this article, we carefully study the frequency features of major interference signals and apply an efficient filter to remove the signals due to non-brain activities. Meanwhile, we propose a sensor fusion–based approach with the help of IMU readings to detect electrodes drifting and assist estimating signal quality.

Second, the high computational complexity of existing emotion recognition algorithms prohibits them to be adopted directly on wearable devices. To address such an issue, we split the emotion recognition into two phases and install them on smart glasses and private clouds (personal PC or smartphone). Instead of extracting emotions in real time, we trigger lifelogging by analyzing the emotional changes. The exact emotion information are obtained offline and tagged to the lifelogs.

Third, the environment impacts the lifelogging qualities. For instance, poor light conditions and high noise levels might influence the qualities of image lifelogs and audio lifelogs, respectively. Meanwhile lifelogging methods, image, audio, and video with various configurations (e.g. resolution and frame rate) have different energy profiles. To best balance the logging quality and the energy consumption, we propose the optimization framework, which picks and configures the suitable lifelogging method with the analysis of both the environment and system context.

In summary, this article makes the following contributions:

- The proposal of a new and natural way to automatically trigger lifelogging. We introduce the use of EEG for the lifelogging service design on wearable devices.
- A series of techniques to integrate EEG electrodes with wearables. We present our signal processing and two-phase emotion recognition design, to meet the computation and energy constraints of wearable-class hardware. We also discuss the optimization framework that selects and configures the suitable lifelogging method to balance lifelog qualities and energy consumptions.
- The full design and implementation of Memento—an emotion-driven lifelogging service on smart glasses. We show Memento is efficient in support of emotion tagging and lifelogging workload. The experimental evaluation indicates that Memento is able to provide satisfactory battery life and user experience.

The rest of this article is organized as follows. Section 2 presents the background of lifelogging and emotion recognition as well as the motivation of our system. Section 3 describes our system design. In Section 4, we introduce the implementation of our prototype system. Section 5 presents the experimental evaluation results. Section 6 discusses the related work, and Section 7 concludes this article.

2 LIFELOGGING AND EMOTIONS

Before we detail the Memento system design, we first elaborate on some preliminary information related to Memento in the following.

Lifelogging. Lifelogging is a form of the digitalization of people's daily, personal experiences. In past decades, several dedicated passive lifelogging devices are proposed for various purposes, e.g., augmenting lives [44], memory aid [22], and health care [34]. With the rapid growth of mobile devices, lifelogging services are gradually migrated to smartphones. Multiple on-board sensors and various sensing techniques enrich the contents of lifelogs, including geo-location information [27], inferred user activities [49], and so on. Social Network Services (SNS) further turns the lifelogging into an emerging lifestyle, and smartphone users love to collect memorial moments and share with friends.

Recently, wearables gain their popularity and show great potential benefits for lifelogging. However, rather than a continuous logging with manual interventions, we believe that lifelogging could be improved by capturing the user's emotions—the system should be able to automatically log memorial moments by understanding the user's mental state changes. For example, when the user gets very happy or sad, the happiness or sadness could be recorded in the forms of audios, images or videos automatically. Some unexpected situations or accidents could be logged as well when the user suffers strong emotional disturbances. We also believe that the lifelogs with emotion information could not only improve the lifelogging system itself, but also enhance many other existing services such SNS, health care, memory aid, and so on. For instance, lifelogging services are used to provide memory cues for the people who struggle with Alzheimer's disease. However lifelogging technologies usually collect an overwhelmingly large amount of lifelogs to review. With the help of intelligent emotion sensing, better memory cues could be selected and recorded automatically, which accord with the people's mental states.

Emotion. Generally speaking, the bodily changes follow directly the perception of the exciting fact, and that our feeling of the changes when they occur is the emotion [25]. Although there are some arguments to explain how emotions occur, the consensus is that emotions are physiological and measurable. Some basic emotions can be found across individuals when certain stimulation is given. For instance, prior studies have classified six [18, 53] or eight [47] basic emotions. It is further improved by the bipolar model, in which the arousal dimension (*how energized the experience feels*) and the valence dimension (*how negative or positive the experience feels*) are considered.

The emotion changes can be observed via physiological effects such as heart beats, facial expressions, voices, and brain activity. EEG is a measure of brain activity via the brain's electric signals. To harvest signals, EEG electrodes are attached on the scalp. The emotion changes lead to distinguished patterns in EEG signals from the certain positions. There is an increasing number of EEG-based emotion recognition techniques [23, 41, 43]. However, most of EEG-based emotion recognition algorithms are not designed for the wearable-class platform. To adapt them, the problem we are faced with is algorithm complexity. The influence of the high complexity is twofold: time constraint and energy overhead. On one hand, most of the algorithms are proposed for the off-line recognition, even on powerful platforms such as PC and workstations. On the other hand, the energy issue rarely has been considered in previous designs.

3 MEMENTO SYSTEM DESIGN

Figure 2 illustrates the architecture of the Memento design with three key modules: the signal processing module, the lifelog collector, and the emotion recognition module. The signal processing module takes EEG readings as input to detect users' emotional changes, e.g., emotion events. When an emotion event is detected, the lifelog collector starts to conduct the lifelogging procedure. The camera and microphone on smart glasses serve as the major medias to capture the daily life moments in the form of videos, audios or pictures, according to both the environment conditions and the system states. Finally the recognized emotions are tagged on the respective lifelogs.

The EEG electrodes often loosely contact on user's scalp and the sensed EEG signals can be easily polluted by interfering activities, e.g., eye blink, head movement, and so on. To address this issue, in the signal processing module we propose an effective band-pass filter to remove ambient noises, utilize the kernel-based correlation to exclude non-brain activities, and further leverage motion sensors and the rigid feature of the wearable to score the quality of the perceived EEG signals (Section 3.1). With all high-score EEG signals, the lifelog collector performs lifelogging according

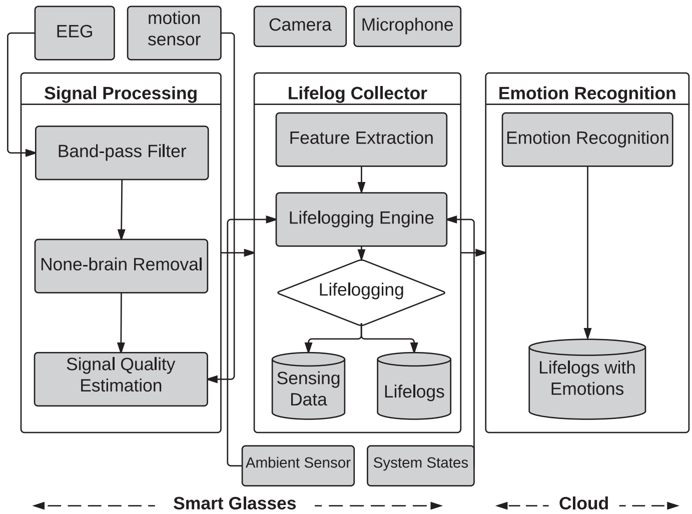


Fig. 2. Architecture of the Memento design.

to user's emotions. However, recognizing emotions from EEG signals is computational intensive, not affordable by wearable platforms directly. To address this issue, we propose a two-phase emotion recognition solution and install them on the smart glasses and the cloud, respectively. On the smart glasses side, without specifying the exact emotion type, Memento merely detects emotion changes by leveraging the intermediate results (Section 3.2). Once a significant change is detected, the lifelogging is launched. To best balance the lifelogging qualities and the energy consumption, we propose the lifelogging engine to automatically select and configure the suitable lifelogging methods, according to the environment contexts and the system states (Section 3.3). Then lifelogging engine would also determine the lifelogging duration dynamically. Finally, the collected lifelogs and sensed data will be uploaded to the cloud later when the smart glasses are recharging or on the user's request. On the cloud, a sophisticated recognition algorithm asynchronously runs to recognize the exact emotion type, which serves as the emotion tag for the subsequent recorded lifelogs.

3.1 Signal Processing

In this section, we first describe the EEG signal processing module in Memento. Although it has been extensively investigated by the studies in the brain-computer-interface domain, they mainly focus on the signals from the EEG electrodes that are firmed attached to the users. However, in the mobile and dynamic scenarios with loosely contacted EEG electrodes in Memento, we need to process the signals with the following steps:

Band-pass filter design. For the collected EEG signals from the headband, we first apply a band-pass filter due to the following reason. The frequency of EEG signals is in the range from 1 to 100Hz, which can be further divided into multiple bands: Alpha (8–14Hz), Beta (12–30Hz), Gamma (>30Hz) and Delta (<4Hz), and so on. As the Alpha and Beta two bands are considered to most related to human's emotion states [12], we thus introduce a 7- to 31Hz band-pass filter (with a 1Hz margin) to extract the EEG signals from these two bands only. With this band-pass filter, we can also filter out two typical interference signals: (1) the heart beat, which is usually from 1 to 6Hz [40], and (2) the electronic interference from nearby appliances, which is at 50Hz (as they are powered by the alternative current whose frequency is 50Hz), as shown in Figure 3. We note that

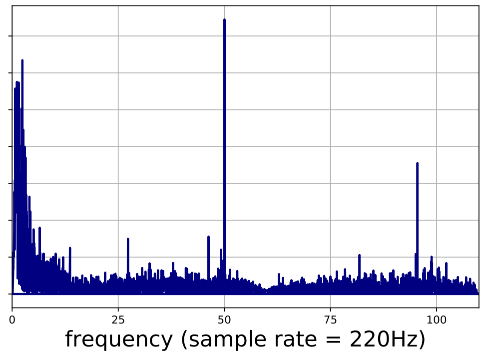


Fig. 3. Frequency distribution of raw EEG signals.

it is possible that there exists some noises that we may not identify yet. However, the bandwidth of our band-pass filter is very small, i.e., 7–31Hz. As long as they are not within this range, they will not affect the Memento's performance.

Non-brain activity removal. Non-brain activities, including chews, blinks, and other muscle activities, produce additional electric signals and can be harvested by EEG electrodes as well. Frequent non-brain activities involve lots of interference to EEG signals. In EEG study, electrodes are usually attached on certain positions such as TP9, TP10, AF7, and AF8 (see Figure 8(c) as a reference) based on the 10-20 system [6]. In this setting, the most significant non-brain activity left is blinks.

As the blink signals have an overlap with the Alpha and Beta bands, which thus cannot be removed by the band-pass filter. To detect and further remove its impact, we notice that the eye blink can be easily detected from the time domain on AF7 and AF8 two channels, as shown in Figure 4. Therefore, once we detect the signals with the user's eye blink, we can apply existing algorithm [36] to remove its impact. To this end, we adopt a time window with length 400ms (corresponding to the common duration of the eye blink) and apply the template matching [52] by using the kernel correlation—a template is correlated with the incoming signals. Once the eye blink is encountered, we can observe a high correlation value and apply the algorithm from Reference [36] in this time window to remove the impact of the eye blink.

Against loosely contacted EEG electrodes. Finally, we consider the signal quality impacted by electrodes' drift. The electrodes of the wearable EEG are not fixed on the scalp. Instead, they are only loosely contacting. However, in the context of lifelogging, the user may move. In this case, because of the loose contact, the electrodes could drift on the scalp. Such a drift actually highlights the design issue. In particular, when the electrode drift occurs, we can also observe the EEG signal changes. We thus need to design an effective way to distinguish the observed EEG signal changes are due to electrodes' drift or user's actual emotions.

To address this issue, our key observation is that when EEG electrodes drift, all the electrodes should be influenced at the same time as the wearable device has a rigid body (i.e., all the electrodes will drift at the same time). On the contrary, the user's actual emotions may cause different EEG signal changes on different electrodes. The cross-correlation mainly serves as a concrete way to quantify the similarity of the signal changes from different electrodes.

- In Memento, we mainly use the AF7 and AF8 two electrodes, and Figure 5 illustrates an example when the drift occurs. To quantify these similarity, we denote R and G as the signal strength of each EEG sample collected from two electrodes, respectively, where $R = \{r_1, r_2, \ldots r_n\}$, $G = \{g_1, g_2, \ldots g_n\}$ and the window size is n. The correlation c between R

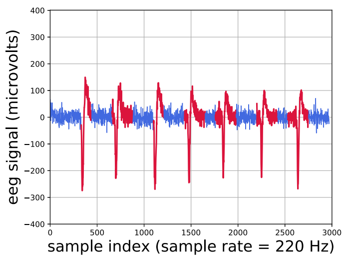


Fig. 4. Detecting blinks from the EEG signals using the kernel correlation.

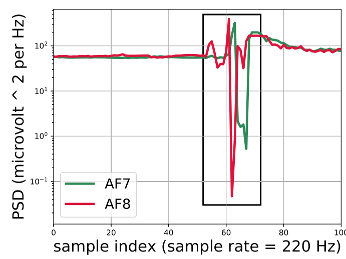


Fig. 5. The EEG signals from electrodes AF7 and AF8 when the drift occurs.

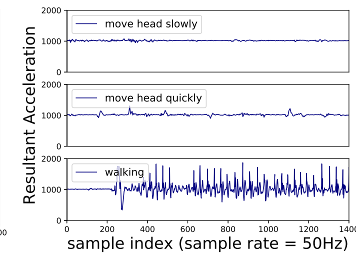


Fig. 6. Resultant accelerometer readings under different mobility levels.

and G is calculated as

$$c = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{r_i - \overline{R}}{S_R}\right)\left(\frac{g_i - \overline{G}}{S_G}\right),$$

where $\overline{R}$ and $\overline{G}$ indicate the average values of R and G, respectively, and S_R and S_G represent their deviations, respectively. The value of c is always between -1 and 1. The larger c indicates two signals are more similar to each other, implying the drift is more likely to occur.

- However, the user's mobility itself will not directly cause the EEG signal changes. However, as the electrodes are not firmed attached to the user's scalp, the mobility could impact their contacting firmness, which in turn may degrade the signal quality (the electrodes may not necessarily have an obvious drift, while a weaken contact could decrease the signal to noise ratio of the signal. In the case where the electrodes indeed drift due to the mobility, its impact will be captured by the correlation stated above). In Figure 6, we find that we can further leverage the motion sensors from the wearable device to infer the mobility levels of the user. For instance, when the user is static or only has the head movement, the variance of the accelerometer readings is small. However, when the user starts to walk, the variance becomes much larger. Therefore, we utilize the variance of the motion sensor data as another indicator of the contact quality $v = Var(Acc)$. Finally, we combine the power correlation c and the acceleration variance v as the estimated signal quality score: $s = \frac{1}{2^{(c+v)}}$.

We can then utilize s to select the good-quality EEG signals and drop the low-quality signals, and the selected good-quality signals will be passed to other component for a further processing.

3.2 Two-phase Emotion Recognition

Emotion recognition algorithms are heavy to be adopted on wearables directly, especially when we consider the stringent real-time requirement. To best our knowledge, most of algorithms are designed for recognizing emotions off-line. For instance, according to our measurement, the algorithm [41] costs about $2.8 \times t$s to perform the emotion recognition only once on the powerful LG Nexus 5X (with the ts EEG signal as input), which clearly cannot satisfy the real-time requirement on the wearable (Google Glass) directly.

Emotion recognition algorithms are based on different theories, but the procedures of these algorithms share the similar principle: They can be decomposed into the feature extraction and classification two phases. We observe that the classification phase is the major contributor to slow down the execution time, while the feature extraction phase is relatively lightweight. This is because most feature extraction approaches leverage near linear algorithms, but plenty of

ALGORITHM 1: Segmentation Strategy

```
segmentation(data);
Input: The EEG data after preprocessing data
Output: Circular buffer
currentSegment ← capture data from startLoc with windowSize;
isValid ← valid(currentSegment);
if isValid then
    index = valid segment nums / bufferSize;
    Buffer[index] ← currentSegment;
end
startLoc ← advance(startLoc, isValid);
return Buffer;

advance(startLoc, isValid);
Input: Last segment start index startLoc and whether it is valid isValid
Output: The start index for the next segment
if isValid then
    consecutiveDrop ← 0;
else
    consecutiveDrop ← consecutiveDrop + 1;
end
startLoc = startLoc + windowSize * (1 − max(0, overlap − consecutiveDrop * 0.1));
return startLoc;

valid(segment);
Input: The current segment segment
Output: If the current segment is accepted isValid
qualityScore ← the signal quality score;
if qualityScore >= threshold − tolerance then
    threshold ← average score of all valid segment;
    return True;
else
    return False;
end
```

iterations are involved in the classification phase. For instance, in Reference [43], the lifting-based wavelet transformation is used to extract features, which is acceptable yet, but in the classification phase, the Fuzzy C-Means clustering is applied, introducing a vast number of computation. Based on this observation, we make an attempt to separate the emotion recognition algorithm into the feature extraction phase and the classification phase, then install them on wearable and cloud, respectively.

EEG segmentation. Before introducing the feature extraction phase on Memento, we first describe our data segmentation strategy. Algorithm 1 shows the pseudo-code of out segment strategy. The design goal is to balance the data redundancy and the processing efficiency. We apply a sliding window of 10s with 2200 samples initially and set 50% overlaps. With a larger window, the emotion recognition algorithm could be feed by more comprehensive data and might provide more accurate estimation, but consumes much more time to process and lead to the greater delay between the certain event occurs and the lifelogging process is triggered. Since Memento is designed as a near

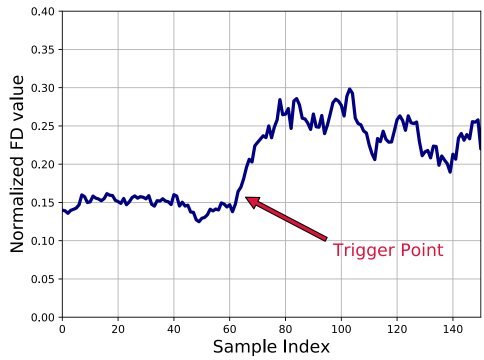


Fig. 7. Using the arousal trend to trigger the lifelogging.

real-time lifelogging service, we initially set a relatively small window and dynamically adjust the overlap size.

According to the signal quality score s stated above, we can determine whether to keep the current segment or drop it. We use an adaptive threshold instead of a fixed one (line 23). However, if the consecutive segments are invalid, it indicates the current situation gets worse. To avoid the unnecessary processing overhead, We reduce the window overlap 10% for each consecutive dropped segment. We always reset the overlap to 50% when a valid segment is found (line 11–18). Finally, we store the valid segments in the circular buffer. The buffer follows the FIFO rule (line 4–6). If there is no space for the coming segment, then the oldest existing one will be replaced.

Feature extraction. From the accepted segments, Memento fetches the EEG segments from the buffer continuously and then extracts features. Since the exact emotions would not be recognized at this phase yet, we consider to use the intermediate results from features to trigger lifelogging. Although these features do not lead to the exact emotion, they can indicate that the user already demonstrates remarkable emotional changes, which can confidently trigger the lifelogging (the emotion tag can be supplemented later after the emotion recognition is completed on the cloud).

In Memento, the selected feature is called fractal dimension (FD), which has been widely used in prior emotion recognition designs [26, 41, 53]. FD is an index for characterizing fractal sets by quantifying their complexity and it is considered to be positively correlated to how energized the user is, hence the arousal level. In particular, we make use of Katz's FD calculation [28] and apply it directly on the waveforms. We obtain the FD of a signal sequence using:

$$D = \frac{\log n}{\log \frac{d}{L} + \log n}, \tag{1}$$

where L is the total length of the signal sequence and d is the diameter that can be estimated as the farthest distance with the beginning point. The parameter n is the number of the steps in the sequence, which is adjusted by the granularity. In particular, we define $n = L/a$, where a is the average distance among the successive points.

Lager FD values are associated with higher arousal levels. To tolerate the absolute arousal differences among individuals, we do not use the absolute arousal level as the threshold to trigger the lifelogging procedure. We propose to use the trend of emotional changes, which is to calculate the derivation of the FD values. The FD trends represents the mental change direction, positive or negative, and also the arousal level. Figure 7 shows the occurrence of a positive trend, and lifelogging is thus triggered after the sample of index 60.

Table 1. Power Consumption and Estimated Battery Life for Different Lifelogging Methods on the Google Glass

Lifelogging Methods	Power Draw	Battery Life
Idle	18mW	116 hours
Audio recording	642mW	3.2 hours
Image capturing	2842mW	0.8 hours
Video capturing	3008mW	0.7 hours

Two-phase emotion classification. Finally, when Memento determines to launch lifelogging, the lifelogs, the EEG features and related sensor readings, including motion sensor data and GPS, will be compressed and saved in the storage of the smart glass. We notice that although the computation and energy resources are limited on wearables, the storage is sufficient. Then the stored data can be uploaded to the private cloud (PC or smartphone) later when the smart glass is recharging or on the user's requests. On the cloud, we run the sophisticated emotion recognition algorithm to recognize the emotions [41].

Memento currently can recognize six emotions, including fear, frustrated, sad, satisfied, pleasant and happy. Such a division is based on References [41, 51]. However, this emotion classification may not be universal. The classifications in different research studies could be slightly different. For instance, Plutchik et al. [47] propose eight emotion classifications, which are anger, fear, anticipation, surprise, joy, sadness, trust, and disgust. As there is no universal emotion classification yet, our current design is mainly based on the emotion classification from References [41, 51, 53].

3.3 Lifelogging Engine

When the lifelogging is triggered, a suitable lifelogging format (by video, audio, or image) should be selected and also properly configured by adjusting the settings such as FPS and resolution, according to the environment conditions and the system states (Section 3.3). The lifelogging duration for video and audio recording each time needs to be decided as well in the lifelogging engine.

Generally speaking, a camera and a microphone on wearables can serve as the major media and lifelogs can be recorded in the forms of video, audio or pictures. In ideal cases, the video clips contain almost all the details of audios and pictures. However, in practice, the environment usually impacts lifelogging qualities. For instance, poor light conditions lead to low-quality videos. Vigorous motion might make the camera go out of focus, in turn leading to blurred pictures. However, each lifelogging formate with various settings, e.g., resolution, FPS, and sample rate, has different energy profiles. Table 1 shows the power consumption and the estimated battery life for audio recording, image capturing, and video capturing on Google Glass. Using expensive settings naturally consumes more energy [38]. For instance, the energy consumption of video capturing is almost 5 times that of audio recording. Therefore it is crucial to balance the lifelogging qualities and the energy consumption.

To tackle this problem, we design the lifelogging engine, which automatically selects and configures the suitable lifelogging method with the analysis of the environment conditions and the system states. By sensing the current environment conditions, including light conditions, movements, and noise levels, we infer the expected qualities of each lifelogging method and formulate it as *Utility*. We then formulate the *Cost* of each lifelogging method in terms of the energy consumption and the resource usage. We also consider the device surface temperature to ensure the user-comfort on wearables. The lifelogging engine thus selects the proper lifelogging format by minimizing the *Cost* and maximizing the *Utility* and then utilizes this selected format to conduct the lifelogging.

Utility. We design the utility function to represent the expected quality of a lifelog. The function assess the utility based on two criteria: how poor the environment conditions are (u_e) and what lifelogging configurations are used (u_c). We apply a linear combination of u_c and the reciprocal of u_e:

$$u^m = w_e^m (u_e^m)^{-1} + w_c^m u_c^m, \tag{2}$$

where $m \in \{\underline{v}ideo, \underline{a}udio, \underline{p}hoto\}$ represents each lifelogging format, respectively, and $w_e^m + w_c^m = 1$.

(1) Calculating u_e^m. To obtain u_e^m, we consider three types of environmental factors: the light condition, the mobility level, and the noise level. We observe each lifelogging format is able to tolerate one or few negative impacts from the environment. For instance, the light condition cannot impact the audio quality, and the noise level cannot influence the photo quality. We thus formulate u_e^m for each lifelogging format as follows:

$$\begin{aligned} u_e^v &= w_l^v \theta_l + w_m^v \theta_m + w_n^v \theta_n, \\ u_e^a &= w_n^a \theta_n, \\ u_e^p &= w_l^p \theta_l + w_m^p \theta_m, \end{aligned} \tag{3}$$

where $\theta_i, i \in \{\underline{l}ight, \underline{m}ovement, \underline{n}oise\}$ stands for current environment conditions and each w_i^m is the weight.

We sense the light condition θ_l via the on-board ambient light sensor, which provides the light intensity in luminance. We extract the movement level θ_m from the accelerometer readings. We borrow the results from the movement estimator in Section 3.1. Then we adapt and simplify the algorithm from Reference [35] to identify the movement state into one of {stationary, walk, ride}. To obtain the acoustic noise level, a one-second acoustic sample is collected. From the sample, we extract the sound pressure level (SPL) as θ_n by using the built-in API.

(2) Calculating u_c^m. To obtain u_c^m, we consider four types of configurations: resolution, frame rate, sampling rate and acoustic channel. From the technical specifications of the device, we can get the "highest" configuration that can be supported by the device for each lifelogging format m, e.g., the highest video resolution of the camera, denoted as $best\_config^m$. Then we can compute u_c^m as

$$u_c^m = current\_config^m / best\_config^m. \tag{4}$$

For the weights used in the *Utility* above, we investigate their settings in Section 5.3.

Cost and system capacities. We assess *Cost* from three aspects: the energy consumption c_e^m, the resource usage c_r^m, and the temperature increment c_t^m:

$$c^m = w_e c_e^m + w_r c_r^m + w_t c_t^m, \tag{5}$$

where $m \in \{\underline{v}ideo, \underline{a}udio, \underline{p}hoto\}$ stands for the lifelogging format. w_e, w_r, and w_t is the particular weights for each cost component. In the current implementation, we set three weights equal to each other.

(1) Calculating c_e^m. We model the energy consumption of each lifelogging format c_e^m as a linear function of the lifelogging duration t, i.e., $c_e^m = p^m t$, where p^m is the power of lifelogging. We experimentally obtain p^m via the off-line tests on wearables. The results on Google Glass are shown in Table 1. For the photo capturing, we set p^p as the average power when taking photos consecutively. The capturing interval is fixed and set to 5s.

(2) Calculating c_r^m. We calculate the resource usage c_r^m by considering CPU and memory utilization of each lifelogging format. We set them as the constants C_{cpu} and C_{memory}, which can be obtained by one-time off-line training.

Table 2. Pre-defined Lifelogging Configurations

Method	Item	Settings		
Video	Resolution	320*180	640*360	1280*720
	Frame Rate	15 FPS	30 FPS	30 FPS
Audio	Sampling Rate	16kHz	22kHz	44kHz
	Channel	Mono	Stereo	Stereo
Image	Resolution	640*360	1280*720	1920*1080

(3) Calculating c_t^m. Usually smart glass directly touches the skin of users. The high device temperature leads to the degraded user-comfort. Thus we also consider the device temperature of wearables and ensure it will not exceed the certain constraint. Since CPU is the major heat contributor, we leverage the CPU load to predict the temperature increment over the duration t.

After we determine $Cost$, the $Capacity$ used in the optimization, in terms of battery budge cap_e, CPU and memory utilization cap_u and device temperature cap_t, are further set as follows:

$$\begin{aligned} cap_b &= \theta_b Budget, \\ cap_u &= \theta_u Memory + \theta_u CPU, \\ cap_t &= T_{max}, \end{aligned} \tag{6}$$

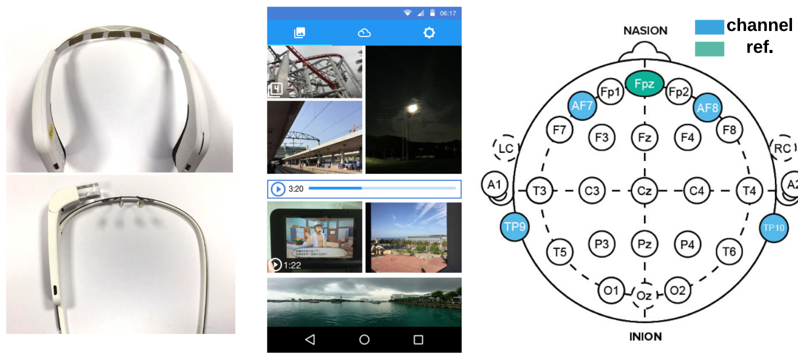
where we empirically set the energy budget constraint θ_b as 90% of the full battery life and resource utilization constraint θ_u as 95% of the memory quota. We set the temperature constraint T_{max} as 48°C to ensure the user-comfort according to Reference [39]. We inquire the current battery life, the available resource quota and CPU temperature by system APIs.

Optimization framework. We define *Utility* and $Cost$, as well as *Capacity*. To select the suitable lifelogging method, we aim to maximize the utility (Equation (2)) by controlling the cost (Equation (5)) less than the capacity (Equation (6)). We regard it as a combinatorial optimization problem, specifically, the 0-1 knapsack problem [50]. Each lifelogging format with certain settings has the estimated utility and the costs in terms of the energy consumption, the resource usage and the heat increment. The *knapsacks* represent the current capacities, which are the battery budget, the available CPU and memory quota and the device temperature. After solving this optimization problem, we get a suitable lifelogging format with the specific setting as well as the lifelogging duration for video and audio capturing. To reduce the optimization search space for wearable devices, we provide some predefined configurations to be used, as in Table 2. We also provide a set of predefined lifelogging duration to speed up, which is $t = 30 + i \cdot 60$s and for i, we set its range from 0 to 9.

3.4 Computation Offloading

All Memento components described so far can be executed purely on wearables. In this section, we introduce a design that can offload certain computations to user's smart phone (if available) for further improving the battery life of wearables.

In particular, we install the same programs and algorithms of the signal processing and emotion event detection. Once the smartphone is available, all the sensed data, including EEG data and motion sensor readings, are processed and analyzed on the phone until a certain emotion event is detected. Then the smartphone in turn notifies the smart glass to trigger the lifelogging engine. We switch the computation between the smart glasses and the smartphone according to three criteria as follows:



(a) Hardware used in the prototype. (b) UI of Memento on the smartphone. (c) Electrodes' contacting spots on the user's scalp.

Fig. 8. The hardware and software implementation of the Memento prototype.

(1) When the smart phone CPU is awake for a period of time, we offload tasks to the smartphone due to the Piggyback effect [30]. We try to avoid waking the smartphone from the idle state to perform tasks, because the tasks are simple though, the CPU and many related subsystems must be activated, which consume can more energy. We piggyback the tasks on the smartphone to maximize the CPU utilization and save the battery life.
(2) When the battery budget on the smart glasses is lower than a threshold, we offload tasks to the smartphone, since we try to use to record as many lifelog as possible before the battery dies. Memento allows the user to set the battery threshold, and by default we set it as 15% of the full battery life.
(3) When the smartphone is busy, we return the offloaded tasks to the smart glasses to ensure the responsiveness. We monitor the circular buffer mentioned in Section 3.2. The time difference between the oldest and newest segments in the buffer indicates if the data are processed in time. We check the time difference when fetching the segment and send tasks back to the smart glasses if it is longer than 240s in our implementation.

4 IMPLEMENTATION

Based on the design in Section 3, we implement a Memento system prototype in this article.

Hardware. As illustrated in Figure 8(a), the current prototype is composed of two commercial devices: a Muse EEG headband [8] and a Google Glass Explorer Edition.

The Muse headband is designed as a personal meditation assistant, which helps the user with meditation exercises. Four channels are supported on the Muse headband, and they are illustrated by the 10-20 system [6]. This headband has five touching spots on the user's scalp, including one reference point in the middle, as shown in Figure 8(c), and other four are EEG measurement electrodes. The electrodes measure the voltage fluctuations resulting from the ionic current within the neurons of the brain on the scalp. In addition to the electrodes, the Muse headband is also equipped with an accelerometer. A PIC24 MCU is used to conduct simple processing operations, e.g., low-pass filtering and band filtering. The headband has a separated battery of 250mAh.

We use a Google Glass as the main wearable platform with Android OS. The core component of a Google Glass is the OMAP4430 system-on-chip (Soc), which owns a dual-core ARM Cortex-A9 CPU, 2BG RAM, and 16GB flash storage. Its rich sensors satisfy the requirements of lifelogging and sensing, including microphone, camera, accelerometer, gyroscope, ambient light sensor, and so on. The Google Glass has a 570mAh liPo battery. In addition, both Wi-Fi and Bluetooth are supported.

In our prototype, the raw EEG reading are preprocessed by the band-pass filter on the Muse headband. Since the headband is not programmable, the rest of the processing operations are performed on Google Glass. EEG data are pushed from the headband to the glasses via Bluetooth. On receiving the EEG data, Google Glass attaches a time stamp to the data. For all other types of data, such as motion data, GPS, and recorded video, they are generated by Google Glass, and we attach time stamp as well when they are generated. Therefore, all the data used in the Memento design can be easily aligned. The resolution of this alignment is high, which is in the level of milliseconds. Finally, the computation tasks of Google Glass can be offloaded to an LG Nexus 5. The back-end service runs on a workstation with a Intel Xeon processor, 16GB RAM, and 1T disk.

Software. We implement the Memento modules in three software packages and deploy them on the Google Glass, the smart phone, and the back-end server, respectively.

- We implement preprocessing and lifelogging two functions as an Android Wear application. Most of the codes are written in Java. For the common utility algorithms used in the data cleaning and feature extracting, e.g., Goertzel algorithm, Katz's FD calculation, and so on, we implement them as the executed libraries in native code.
- Memento service on the smartphone includes two major features. First, we build the offloading host as a standard Android service. The service remains in sleep until certain offloading requests are received through the notification mechanism. Since the smartphone runs Android as well, the services share the same set of native libraries and the sub-set of Java code with the Memento application on Google Glass. We use a Wi-Fi Ad hoc network to set up the connection between the smartphone and the glasses. Second, we implement the application with GUI to browse the lifelogs in the private cloud and the Google Glass. The application is also used to share the lifelogs to the SNS easily. Figure 8(b) shows the current user interface on Android.
- Back-end server. We implement the full version of the emotion recognition on the back-end server in Java. We setup a web service on the server. Lifelogs along with related sensing data uploaded through the encrypted links. The back-end sever also provides the features of querying and browsing lifelogs.

Lifelog Store. Lifelogs are distributed both on wearables and the private cloud. On the Google Glass, media lifelogs are stored in the raw format. Before storing the EEG features, we compress them to reduce the size. We use SqLite database to maintain the index. An entry in the index represents a lifelogging event, which contains the corresponding media lifelogs and EEG data. Currently, we have four attributions in each entry, which are the timestamp, the location obtained from GPS on the smartphone, the file pointers of the media lifelogs, and the EEG features. The lifelogs on the Google Glass are uploaded to the private cloud when the glasses are in charge or on the user's request. On the private cloud, lifelogs are stored in the similar structure with on the glasses, but the emotion tags are added. Each entry is further assigned one type of basic emotions, along with its arousal and valence levels.

5 EVALUATION

In this section, we present our evaluation on Memento. We first introduce the experimental setting and then report the detailed system performance of Memento.

5.1 Experimental Setup

(1) Related to emotion recognition. In this experiment, we adopt two datasets: a public dataset DEAP [29] with 32 participants and our collected dataset with five participants (their information is summarized in Table 3). We are given the access to the DEAP dataset, which is collected

Table 3. The Information of Our Recruited Five Volunteers

ID	Age	Gender	Handedness	Vision	Vision Aid	Education	Alcohol	Coffee	Tea	Tobacco	Hours of sleep
C01	28	Male	Right	Corrected to normal	Glasses	BA	Never	Regularly	Never	Never	8
C02	27	Male	Right	Normal	No	BA	Never	Never	Regularly	Never	6–8
C03	30	Male	Right	Correct to normal	Glasses	Ph.D	Never	Never	Regularly	Regularly	5–8
C04	26	Female	Right	Correct to normal	Glasses	BA	Never	Never	Never	Never	8
C05	24	Male	Right	Normal	No	BA	Never	Never	Never	Never	7–9

for the analysis of human affective states. In DEAP, the EEG signals of 32 participants are collected by a dedicated EEG equipment with 32 electrode channels of a good signal quality, while each participant watches 40 one-minute-long excerpts of videos. After they watch one video clip, a self-assessment tool [13] is used to let participants rate their reflections for each video in terms of the level (1 to 9) of arousal, valence, and dominance. The size of DEAP dataset is 2.7GB in total. As the dataset DEAP has a clear labelling in the arousal and valence two-dimensional (2D) plane, we have conducted the following comparison with the recent emotion recognition design [41] as follows.

- **DEAP-Manual:** For each 1-minute video clip, as all the participants also manually label in the arousal and valence two-dimensional plane, we can recognize their emotions based on such labels using Reference [41].
- **DEAP-EEG:** For each 1-minute video clip, the high-quality EEG data provided by the DEAP dataset (for each participant) can be used to derive a series of points in the arousal and valence two-dimensional plane, based on which we can also recognize each participant's emotion when he/she watched the video using the algorithm proposed in [41].
- **Memento:** For each 1-minute video clip, we let each of our five participants watch the video and record their EEG signals using MUSE headband (the EEG data quality is not as good as that in the DEAP dataset as the EEG data in DEAP is collected by a dedicated equipment). We then apply Memento on our collected data for the emotion recognition.

To compare the performance, we adopt the metric of *Manhattan distance* (please note that we cannot directly compare the emotion recognition accuracy, because the dataset does not label each participant's emotion ground truth). For this metric, the smaller the Manhattan distance is, the closer the compared two methods perform.

(2) Related to the overall lifelogging. We let five participants wear our Memento system, and we collect their EEG readings, motion sensor readings and lifelogs under different scenarios, including laboratory, office, street, and park. When the significant emotion changes occur, the lifelogging collector is triggered. The selected lifelogging method along with the particular settings, the timestamp, and the duration are recorded as well. Totally, we get about 6 hours videos, sensor data traces, and corresponding event timestamps. Finally, we also use the Monsoon power monitor to collect the real-time energy consumption traces of the Google Glass and the smartphones to examine the energy consumption of our design.

5.2 Emotion Recognition Performance

Overall performance. According to the Arousal-Valence emotion model, the user's EEG signals are first converted to a set of emotion samples (points) on the two-dimensional Arousal–Valence plane, based on which we can recognize human's general emotions: fear, frustrated, sad, satisfied, pleasant, and happy according to References [51, 53]. As stated above, we train the emotion

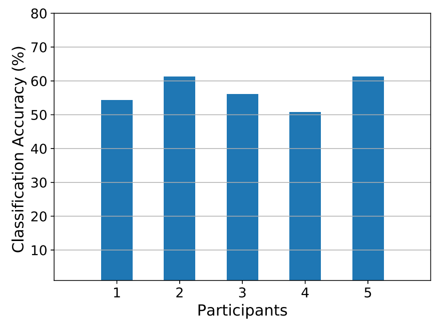
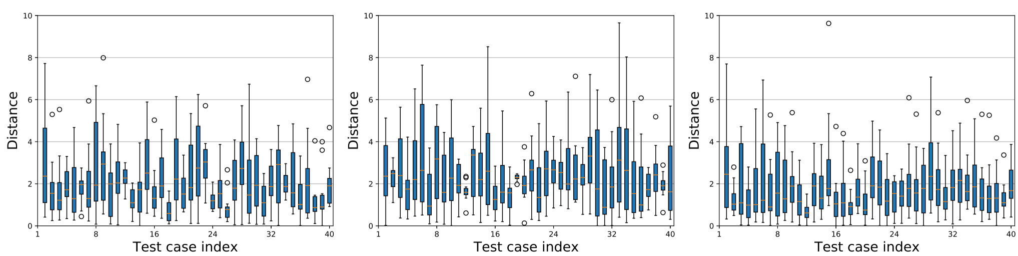


Fig. 9. Overall emotion recognition performance of the five participants.



(a) DEAP-EEG V.S. DEAP-Manual (b) Memento V.S. DEAP-EEG (c) Memento V.S. DEAP-Manual

Fig. 10. Manhattan distance among DEAP-Maunal, DEAP-EEG, and Memento.

recognizer using the dataset DEAP. After the training, we test on these five participants (their data are not used in the training). In particular, after watching one video clip, we let each volunteer select the emotion from the six emotion candidates. Then, based on their EEG data, we further use Memento to recognize their emotions and compare with the volunteers' manually labelled ground truth.[1] The result is as depicted in Figure 9. From the figure, we can see that the emotion recognition of Memento can achieve good performance. The accuracy is from 51% to 61%, and the average accuracy is 57%, which is quite close to the accuracy about 60% using a dedicated EEG equipment in Reference [53]. For the emotion recognition, we leverage existing algorithm from the neurophysiology domain. With the further advancing of this research field, we can update the emotion recognition algorithm for achieving higher accuracy in the future. As an immediate remedy solution against the unreliable emotion recognition labels, the user could review the recognized motion labels and perform a manual correction if needed.

Manhattan distance. To have a detailed understanding of the emotion recognition performance achieved in Figure 9. In Figure 10(a), we directly compare the Manhattan distance of the emotion samples on the Arousal–Valence plane with respect to DEAP-Manual. Figure 10(a) illustrates the distance distribution of all the 40 test cases between DEAP-EEG and DEAP-Manual. As their data are both from the DEAP dataset, we adopt such a distribution as the baseline to understand the performance of Memento. In each test case, we calculate the Root Mean Square (RMS) between DEAP-EEG and DEAP-Manual. Overall the average RMS is 2.42 and the maximal and minimal RMS is 4.25 and 0.90, respectively. Among them, 60% of cases is less than 2.5 and the RMS of over 80% test cases are less than 3.29.

[1]Such an experimental methodology follows prior studies, but it could have the possibility with a prior likelihood of correct guessing by the volunteers.

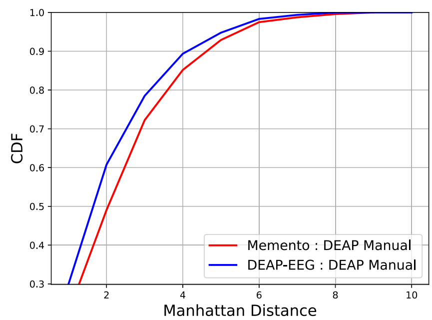


Fig. 11. Manhattan distance distribution for Memento and DEAP-EEG compared with DEAP-Manual.

Table 4. Breakdown of the Emotion Recognition Performance

Process	Averaged RMS
Full design	2.76
Band pass filter (BPF)	3.63
BPF + Non-brain removal	2.92
BPF + Signal quality estimation	3.32

In Figure 10(b), we then compare Memento with DEAP-Manual. Overall, the average RMS is 2.76 and the maximal and minimal RMS are 4.22 and 1.5, respectively. In all the test cases, the RMS of more than 80% cases is less than 3.66. The result indicates that the Manhattan distance increases around 10% merely comparing with the baseline, which explains why Memento can achieve good performance in Figure 9. In Figure 10(c), we also compare Memento with DEAP-EEG. We align the converted emotion samples on the Arousal–Valence plane in the time domain and calculate the distance between each pair of two samples (one from Memento and one from DEAP-EEG) that are closest in time. The results show that the average RMS is 2.12 and RMS from more than 80% cases is less than 2.7. This shows that Memento can perform comparably with DEAP-EEG, which is another indicator that Memento achieves good performance.

Finally, Figure 11 summarizes the detailed Manhattan distance CDFs for both Memento and DEAP-EEG compared with DEAP-Manual, which shows very comparable performance.

Performance gain from different modules. For the emotion recognition of Memento, we propose three modules to improve the EEG signal quality on wearable devices, namely band pass filter (BPF), non-brain activity removal and the countermeasure against the loosely contacted EEG electrodes. In Table 4, we further examine the efficacy of these modules. The average RMS of Memento with respect to DEAP-Manual is 2.76. If we purely adopt the band pass filter, then the average RMS increases to 3.63. If we further add the non-brain activity removal module, then the RMS is decreased to 2.92. The last module can further reduce the RMS value to 2.76. We note that the reduction of the last module is less than the second module is mainly because the occurrence of the wearable device shift is much less frequent than the non-brain activities.

5.3 Lifelogging Performance

We now evaluate the overall lifelogging performance of Memento. In particular, we examine how our emotion-driven lifelogging design can meet the users' expectations.

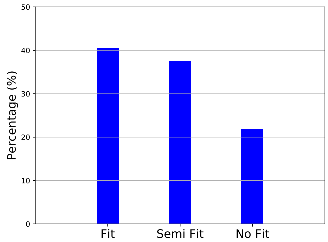


Fig. 12. How the emotion-driven lifelogs fit the users' expectations.

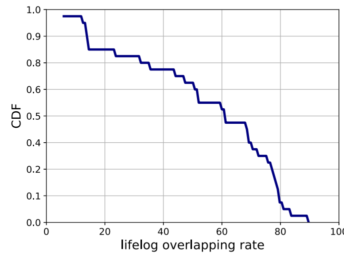


Fig. 13. Coverage ratio of lifelogs against the clips selected by users.

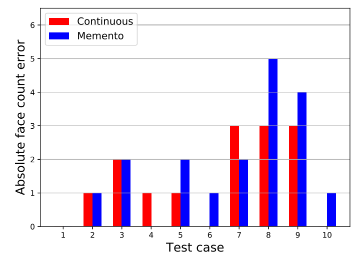


Fig. 14. Lifelog quality evaluation based on the face count.

Fits of user's expectations. In this experiment, we let participants wear our Memento system and we collect their EEG readings, motion sensor readings and lifelogs. For the evaluation purpose, we also store the whole first-person view video (about 6 hours) as the ground truth. After the experiment, we ask each participant to watch its own first-person video and manually select all the clips that they believe should be captured, i.e., by labeling the start and end time. Then we compare the timestamps labeled by participants with the results obtained from Memento. In the comparison, if the video lifelogs are selected by Memento, we define three categories:

- *Fit*: the overlap is more than 50%,
- *Semi-fit*: the overlap ranges from 30% to 50%, and
- *Non-fit*: the overlap is less than 30%.

For audio lifelogs, we define three similar categories but with more strict criteria due to less information contained in the audio lifelogs. We set *fit. semi-fit, none-fit* when the overlap rate more than 65%, 45–65%, and less than 45%, respectively. If the picture lifelog is recorded by Memento, then we regard it as *fit* if it appears in the selected videos by the manual labeling; otherwise, it is *none-fit*. As Figure 12 shows, Memento achieves very good lifelogging performance, where nearly 80% of recorded lifelogs are within the Fit and Semi-Fit two categories.

However, we also evaluate the coverage of the lifelogs selected by participants against the lifelogs collected by Memento. We define the participant selected video is not covered when the time overlap with the Memento lifelogs in any forms is less than 30%; otherwise, it is covered. Figure 13 shows that more than 70% of lifelogs are covered for Memento in the experiment.

Performance of the vision task. To further understand the quality of the collected lifelogs, we utilize one common vision-based task in Memento, i.e., fact detection, to evaluate the performance. With the recorded first-person view videos for each participant, we apply an open-sourced face detection library.[2] We manual label the faces appeared as the ground truth. According to the lifelogging timestamps, we get the corresponding video clips from the first-person video above. We first run the fact detection algorithm on the continuous video as the reference, denoted as Continuous in Figure 14. Then based on the lifelogging decisions that Memento makes (video, audio or photo), we further obtain the number of heads detected from the lifelogs. Figure 14 shows the performance over 10 randomly selected test cases, which shows that Memento achieves comparable performance as Continuous, where the average head counting errors are 58% and 49%, respectively.

However, by using the detection results, we can further get the F1-score [11], which can be used to train the *Utility* and the corresponding weights in Section 3.3.

[2]https://github.com/ageitgey/face_recognition.

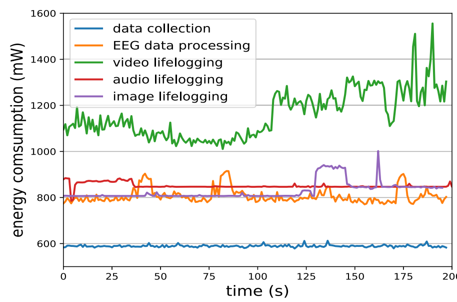


Fig. 15. Energy consumption of each component.

5.4 Energy Consumption

Currently, we implement our prototype via two separate commercial wearable devices: a MUSE EEG headband and a Google Glass. We present their energy consumption in this subsection.

(1) EEG headband. By default the MUSE EEG headset continues to send out the raw EEG data via Bluetooth. In addition, it uses a micro control unit to apply the pre-processing operations on the EEG signals such as smoothing and band-pass filtering. Under this setting, we have measured that the battery life of the EEG headset is about 4.2 hours with the battery size of 250mAh. In the future, we plan to adopt the duty-cycle technique to reduce the sampling frequency, which can further improve the lifetime of the headband.

(2) Google Glass. As Google Glass needs to work in different modes for processing different tasks, we thus measure the detailed energy consumption under each mode in the following.

- Data collection: Memento continuously receives the EEG data from the headband. As shown in Figure 15, data collection contributes to about 589mW energy consumption on average.
- EEG data processing: Memento cleans the data, extracts the FD values, and computes the arousal changes. From the measurement, we find that in addition to the data collection, after Memento processes the EEG data, 200mW more energy will be consumed, as shown in Figure 15.
- Lifelogging: When the arousal changes are detected, the lifelogging is launched. Figure 15 also shows the energy consumption under three types of lifelogging methods. In particular, when audio and image are used, the additional energies are about 40[3] and 153mW, respectively. When video is adopted, the energy consumption could jump to 1200–1600mW.

Different from the headband, currently the co-processor or LPU cannot be utilized on Google Glass, since Android does not release such APIs. According to the current setting, the lifetime of Google Glass is about 4.6 hours. In the future, we envision that the communication cost could be largely reduced or even removed if the EEG electrodes could be physically connected to (or included by) the smart glass device.

6 RELATED WORK

We review existing works that are related to the Memento design from the following two aspects:

Emotion recognition and its applications. Emotion is one of the human natures and can be inferred in several ways [54]. Existing work exploits the recognition of emotions by leveraging the acoustic [32] or the visual signals [9, 16, 45], where the facial or spoken expressions are carefully analyzed. Comparing to the voice and video, the physiological signals are more direct indicators of how we are feeling. The heart beat, breath rates, and blood pressure are used to extract affective

[3]The measured 40mW is the average energy consumption to take the image once.

states in References [20, 55]. Memento makes of EEG. There are a number of EEG-based emotion recognition algorithms proposed in recent years. For example, to classify basic emotions, support vector machine is used in Reference [23]. In Reference [43], wavelet-based methods are applied to extract features, fuzzy k-means and fuzzy c-means clustering are considered to do the classification. Besides the statistic-based methods and wavelet transform-based methods, more approaches are considered. In Reference [46], higher-order crossing analysis is adopted. The authors of Reference [41] make use of the fractal dimension-based algorithm to quantify the basic emotions. However, these algorithms mainly focus the emotion recognition accuracy. The time constraints and the energy consumptions are rarely considered. Most of them are proposed for offline recognition process. The authors of Reference [41] build a real-time emotion recognition system, but it is not affordable for the wearable-class hardware in terms of both the computation capacities and battery life.

Recent years a devise of sensing algorithms on smartphones are proposed to offer a rich set of user signals, including affective activities. Some indirect approaches are explored. EmotionSense [48] builds an audio-based emotion recognition on smartphones. In References [15, 33], the relationship between the phone usage pattern and the personalities is discussed. The authors of MoodScope [37] propose their findings that by analyzing the communication history and application usage patterns, users' daily mood can be inferred accurately. Comparing to these systems, Memento leverages EEG directly and interprets the users' affective activities in the perspective of physiological properties.

Lifelogging system. The origin goal of the lifelogging system is to digitize the users' activities and to provide the query services. In MyLifeBits [19], various activities are collected, including emails, photos, presentation slides, home videos, and audio recordings. It is considered one of the first bite to the comprehensive lifelogging system. In the past decades, many lifelogging systems have been proposed. We categorized them into two groups: dedicated equipment for certain purposes and lifelogging systems on personal mobile devices, smartphones specifically. SenseCam [22] is a body mounted camera, which passively captures photos through a wide-angle lens. SenseCam is proposed for the memory aid. Extending it, Footprint Tracker [21] studies the effects of multiple memory cues. In Reference [34], a ubiquitous lifelogging system is also designed for the episodic memory impairment. Recently wearables gain the popularities. Some wearable cameras [1, 2, 5] can be used to log daily lives. However, most of these lifelogging devices are designed to passively log or request users' intervention. Memento aims to provide a automatic, affordable lifelogging solution on COTS smart glasses.

Lifelogging systems on smartphones are also emerging. Leveraging the rich on-board sensors and sensing algorithms, some context-aware approaches [17, 42] are proposed and can be used to improve lifelogging. Experience Explorer [10] not only senses and captures contextual lifelogs but also provides social network features. Lifelogger [14] collects a diverse of sensor data and focuses on providing the robust lifelog query services. In UbiqLoq [49], authors propose a lightweight framework allowing developers easily create lifelogging application based on it. Recent ZOE [31] leverages the advanced system-on-chip (Soc) techniques. It supports continuously sensing a number of user activities (physical, personal, and social) and provides the dialog-based user interaction. These lifelogging systems are developed to record the external environment. Meanwhile the internal experience of people is important and helpful for lifelogging as well [24]. Memento makes attempts to leverage the user's emotions to improve the quality of lifelogs.

7 CONCLUSION

In this article, we present the design and implementation of Memento, an emotion-driven lifelogging system on wearables. Memento senses the emotional changes of users and automatically launches the lifelogging based on that. So far as we know, Memento is the first-of-its-kind lifelogging system. Through a series of techniques, Memento integrates EEG and proposes the two phase emotion recognition that makes it efficient and affordable on wearables. Finally, Memento outputs lifelogs tagged with emotion information that we believe could enhance many existing services.

REFERENCES

[1] GoPro. 2015. Retrieved from https://gopro.com/.

[2] Vicon. 2015. Retrieved from https://www.vicon.com/.

[3] Everyday. 2016. Retrieved from http://everyday-app.com/.

[4] Instant. 2016. Retrieved from http://instantapp.today/.

[5] Narrative Clip. 2016. Retrieved from http://getnarrative.com/.

[6] 10-20 system. 2017. Retrieved from https://goo.gl/8Yzwv5.

[7] Electroencephalography. 2017 Retrieved from https://goo.gl/okX9pC.

[8] Muse EEG headband. 2017. Retrieved from http://www.choosemuse.com/.

[9] Microsoft Cognitive Services. 2018. Retrieved from https://www.microsoft.com/cognitive-services/en-us/emotion-api.

[10] P. Belimpasakis, K. Roimela, and Y. You. 2009. Experience explorer: A life-logging platform based on mobile context collection. In *Proceedings of the 2009 3rd International Conference on Next Generation Mobile Applications, Services and Technologies*. 77–82.

[11] Christopher M. Bishop. 2006. *Pattern Recognition and Machine Learning*. Springer.

[12] Danny Oude Bos. 2007. EEG-based emotion recognition the influence of visual and auditory stimuli.

[13] Margaret M. Bradley and Peter J. Lang. 1994. Measuring emotion: The self-assessment manikin and the semantic differential. *J. Behav. Ther. Exp. Psychiat.* 25, 1 (1994), 49–59.

[14] Snehal Chennuru, Peng-Wen Chen, Jiang Zhu, and Joy Ying Zhang. 2010. Mobile lifelogger–Recording, indexing, and understanding a mobile user's life. In *Proceedings of the International Conference on Mobile Computing, Applications, and Services*. Springer, 263–281.

[15] Gokul Chittaranjan, Jan Blom, and Daniel Gatica-Perez. 2011. Who's who with big-five: Analyzing and classifying personality traits with smartphones. In *Proceedings of the 2011 15th Annual International Symposium on Wearable Computers (ISWC'11)*. IEEE Computer Society, Los Alamitos, CA, 29–36.

[16] Jeffrey F. Cohn. 2006. Facial expression and emotion. In *Proceedings of the 8th International Conference on Multimodal Interfaces (ICMI'06)*. ACM, New York, NY, 233–238.

[17] Sunny Consolvo, David W. McDonald, Tammy Toscos, Mike Y. Chen, Jon Froehlich, Beverly Harrison, Predrag Klasnja, Anthony LaMarca, Louis LeGrand, Ryan Libby, Ian Smith, and James A. Landay. 2008. Activity sensing in the wild: A field trial of ubifit garden. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'08)*. ACM, New York, NY, 1797–1806.

[18] Paul Ekman. 1992. An argument for basic emotions. *Cogn. Emot.* 6, 3–4 (1992), 169–200.

[19] Jim Gemmell, Gordon Bell, and Roger Lueder. 2006. MyLifeBits: A personal database for everything. *Commun. ACM* 49, 1 (Jan. 2006), 88–95.

[20] A. Gluhak, M. Presser, L. Zhu, S. Esfandiyari, and S. Kupschick. 2007. Towards mood based mobile services and applications. In *Proceedings of the 2Nd European Conference on Smart Sensing and Context (EuroSSC'07)*. Springer-Verlag, Berlin, 159–174.

[21] Rúben Gouveia and Evangelos Karapanos. 2013. Footprint tracker: Supporting diary studies with lifelogging. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'13)*. ACM, New York, NY, 2921–2930.

[22] Steve Hodges, Lyndsay Williams, Emma Berry, Shahram Izadi, James Srinivasan, Alex Butler, Gavin Smyth, Narinder Kapur, and Ken Wood. 2006. SenseCam: A retrospective memory aid. In *Proceedings of the 8th International Conference on Ubiquitous Computing (UbiComp'06)*. 177–193.

[23] Robert Horlings, Dragos Datcu, and Leon J. M. Rothkrantz. 2008. Emotion recognition using brain activity. In *Proceedings of the 9th International Conference on Computer Systems and Technologies and Workshop for PhD Students in Computing (CompSysTech'08)*. ACM, New York, NY, Article 6, 6:II.1–6:1 pages.

[24] Leonid Ivonin, Huang-Ming Chang, Wei Chen, and Matthias Rauterberg. 2013. Unconscious emotions: Quantifying and logging something we are not aware of. *Pers. Ubiq. Comput.* 17, 4 (Apr. 2013), 663–673.

[25] William James. 1884. What is an emotion? *Mind* 9, 34 (1884), 188–205.

[26] R. Jenke, A. Peer, and M. Buss. 2014. Feature extraction and selection for emotion recognition from EEG. *IEEE Trans. Affect. Comput.* 5, 3 (Jul. 2014), 327–339.

[27] Vaiva Kalnikaite, Abigail Sellen, Steve Whittaker, and David Kirk. 2010. Now let me see where I was: Understanding how lifelogs mediate memory. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'10)*. ACM, New York, NY, 2045–2054.

[28] Michael J. Katz. 1988. Fractals and the analysis of waveforms. *Comput. Biol. Med.* 18, 3 (1988), 145–156.

[29] S. Koelstra, C. Muhl, M. Soleymani, J. S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras. 2012. DEAP: A database for emotion analysis using physiological signals. *IEEE Trans. Affect. Comput.* 3, 1 (Jan. 2012), 18–31.

[30] Nicholas D. Lane, Yohan Chon, Lin Zhou, Yongzhe Zhang, Fan Li, Dongwon Kim, Guanzhong Ding, Feng Zhao, and Hojung Cha. 2013. Piggyback crowdsensing (PCS): Energy efficient crowdsourcing of mobile sensor data by exploiting smartphone app opportunities. In *Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems (SenSys'13).* ACM, New York, NY, Article 7, 14 pages.

[31] Nicholas D. Lane, Petko Georgiev, Cecilia Mascolo, and Ying Gao. 2015. ZOE: A cloud-less dialog-enabled continuous sensing wearable exploiting heterogeneous computation. In *Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services (MobiSys'15).* 273–286.

[32] Chul Min Lee and S. S. Narayanan. 2005. Toward detecting emotions in spoken dialogs. *IEEE Trans. Speech Audio Process.* 13, 2 (Mar. 2005), 293–303.

[33] H. Lee, Young Sang Choi, Sunjae Lee, and I. P. Park. 2012. Towards unobtrusive emotion recognition for affective social communication. In *Proceedings of the 2012 IEEE Consumer Communications and Networking Conference (CCNC'12).* 260–264.

[34] Matthew L. Lee and Anind K. Dey. 2008. Lifelogging memory appliance for people with episodic memory impairment. In *Proceedings of the 10th International Conference on Ubiquitous Computing (UbiComp'08).* 44–53.

[35] Jonathan Lester, Tanzeem Choudhury, Nicky Kern, Gaetano Borriello, and Blake Hannaford. 2005. A hybrid discriminative/generative approach for modeling human activities. In *Proceedings of the 19th International Joint Conference on Artificial Intelligence (IJCAI'05).* 766–772.

[36] Yandong Li, Zhongwei Ma, Wenkai Lu, and Yanda Li. [n. d.]. Automatic removal of the eye blink artifact from EEG using an ICA-based template matching approach. *Physiol. Measure.* 27, 4 ([n. d.]), 425.

[37] Robert LiKamWa, Yunxin Liu, Nicholas D. Lane, and Lin Zhong. 2013. MoodScope: Building a mood sensor from smartphone usage patterns. In *Proceeding of the 11th Annual International Conference on Mobile Systems, Applications, and Services (MobiSys'13).* ACM, New York, NY, 465–466.

[38] Robert LiKamWa, Bodhi Priyantha, Matthai Philipose, Lin Zhong, and Paramvir Bahl. 2013. Energy characterization and optimization of image sensing toward continuous mobile vision. In *Proceeding of the 11th Annual International Conference on Mobile Systems, Applications, and Services (MobiSys'13).* ACM, New York, NY, 69–82.

[39] Robert LiKamWa, Zhen Wang, Aaron Carroll, Felix Xiaozhu Lin, and Lin Zhong. 2014. Draining our glass: An energy and heat characterization of google glass. In *Proceedings of 5th Asia-Pacific Workshop on Systems (APSys'14).* 10:1–10:7.

[40] Chia-Hung Lin. 2008. Frequency-domain features for ECG beat discrimination using grey relational analysis-based classifier. *Comput. Math. Appl.* 55, 4 (Feb. 2008), 680–690.

[41] Yisi Liu, Olga Sourina, and Minh Khoa Nguyen. 2011. Real-time EEG-based emotion recognition and its applications. In *Transactions on Computational Science XII.* 256–277.

[42] Hong Lu, Jun Yang, Zhigang Liu, Nicholas D. Lane, Tanzeem Choudhury, and Andrew T. Campbell. 2010. The jigsaw continuous sensing engine for mobile phone applications. In *Proceedings of the 8th ACM Conference on Embedded Networked Sensor Systems (SenSys'10).* ACM, New York, NY, 71–84.

[43] M. Murugappan, M. Rizon, R. Nagarajan, S. Yaacob, I. Zunaidi, and D. Hazry. 2008. Lifting scheme for human emotion recognition using EEG. In *Proceedings of the 2008 International Symposium on Information Technology*, Vol. 2. 1–7.

[44] David H. Nguyen, Gabriela Marcu, Gillian R. Hayes, Khai N. Truong, James Scott, Marc Langheinrich, and Christof Roduner. 2009. Encountering SenseCam: Personal recording technologies in everyday life. In *Proceedings of the 11th International Conference on Ubiquitous Computing (UbiComp'09).* 165–174.

[45] Maja Pantic and Leon J. M. Rothkrantz. 2000. Automatic analysis of facial expressions: The state of the art. *IEEE Trans. Pattern Anal. Mach. Intell.* 22, 12 (Dec. 2000), 1424–1445.

[46] P. C. Petrantonakis and L. J. Hadjileontiadis. 2010. Emotion recognition from EEG using higher order crossings. *IEEE Trans. Inf. Technol. Biomed.* 14, 2 (Mar. 2010), 186–197.

[47] Robert Plutchik. 1990. Emotions and psychotherapy: A psychoevolutionary perspective. In *Emotion, Psychopathology, and Psychotherapy.* Elsevier, 3–41.

[48] Kiran K. Rachuri, Mirco Musolesi, Cecilia Mascolo, Peter J. Rentfrow, Chris Longworth, and Andrius Aucinas. 2010. EmotionSense: A mobile phones based adaptive platform for experimental social psychology research. In *Proceedings of the 12th ACM International Conference on Ubiquitous Computing (UbiComp'10).* ACM, New York, NY, 281–290.

[49] Reza Rawassizadeh, Martin Tomitsch, Katarzyna Wac, and A. Min Tjoa. 2013. UbiqLog: A generic mobile phone-based life-log framework. *Pers. Ubiq. Comput.* 17, 4 (Apr. 2013), 621–637.

[50] K. W. Ross and D. Tsang. 1988. The stochastic knapsack problem. In *Proceedings of the 27th IEEE Conference on Decision and Control*, Vol. 1. 632–633.

[51] James A. Russell. 1980. A circumplex model of affect. *J. Pers. Soc. Psychol.* 39, 6 (1980), 1161.

[52] Longfei Shangguan, Zheng Yang, Alex X. Liu, Zimu Zhou, and Yunhao Liu. 2017. STPP: Spatial-temporal phase profiling-based method for relative RFID tag localization. *IEEE/ACM Trans. Netw.* 25, 1 (2017), 596–609.

[53] Olga Sourina and Yisi Liu. 2011. A fractal-based algorithm of emotion recognition from EEG using arousal-valence model. In *Biosignals*. 209–214.

[54] Zhihong Zeng, Maja Pantic, Glenn I. Roisman, and Thomas S. Huang. 2007. A survey of affect recognition methods: Audio, visual and spontaneous expressions. In *Proceedings of the 9th International Conference on Multimodal Interfaces (ICMI'07)*. ACM, New York, NY, 126–133.

[55] Mingmin Zhao, Fadel Adib, and Dina Katabi. 2016. Emotion recognition using wireless signals. In *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking (MobiCom'16)*. ACM, New York, NY, 95–108.

Received April 2018; revised August 2018; accepted September 2018