Discovering Digital Representations for Remembered Episodes from Lifelog Data

Dudzik, Bernd; Broekens, Joost; Neerincx, Mark; Olenick, Jeffrey; Chang, Chu-Hsiang; Kozlowski, Steve W. J.; Hung, Hayley

DOI
10.1145/3279810.3279850

Publication date
2018

Document Version
Final published version

Published in
Proceedings of the Workshop on Modeling Cognitive Processes from Multimodal Data, MCPMD'18

Discovering Digital Representations for Remembered Episodes from Lifelog Data

Bernd Dudzik*
Delft University of Technology
Delft, The Netherlands
B.J.W.Dudzik@tudelft.nl

Joost Broekens
Delft University of Technology
Delft, The Netherlands
D.J.Broekens@tudelft.nl

Mark Neerincx
Delft University of Technology
Delft, The Netherlands
M.A.Neerincx@tudelft.nl

Jeffrey Olenick
Michigan State University
East Lansing, Michigan, United States
olenickj@msu.edu

Chu-Hsiang Chang
Michigan State University
East Lansing, Michigan, United States
cchang@msu.edu

Steve W. J. Kozlowski
Michigan State University
East Lansing, Michigan, United States
stevekoz@msu.edu

Hayley Hung
Delft University of Technology
Delft, The Netherlands
H.Hung@tudelft.nl

ABSTRACT

Combining self-reports in which individuals reflect on their thoughts and feelings (*Experience Samples*) with sensor data collected via ubiquitous monitoring can provide researchers and applications with detailed insights about human behavior and psychology. However, meaningfully associating these two sources of data with each other is difficult: while it is natural for human beings to reflect on their experience in terms of *remembered episodes*, it is an open challenge to retrace this subjective organization in sensor data referencing objective time.

Lifelogging is a specific approach to the ubiquitous monitoring of individuals that can contribute to overcoming this *recollection gap*. It strives to create a comprehensive timeline of semantic annotations that reflect the impressions of the monitored person from his or her own subjective point-of-view.

In this paper, we describe a novel approach for processing such lifelogs to situate remembered experiences in an objective timeline. It involves the computational modeling of individuals' memory processes to estimate segments within a lifelog acting as plausible *digital representations* for their recollections. We report about an empirical investigation in which we use our approach to discover plausible representations for remembered social interactions between participants in a longitudinal study. In particular, we describe an exploration of the behavior displayed by our model for memory processes in this setting. Finally, we explore the representations discovered for this study and discuss insights that might be gained from them.

*This is the corresponding author

MCPMD'18 , October 16, 2018, Boulder, CO, USA

ACM ISBN 978-1-4503-6072-2/18/10…$15.00
https://doi.org/10.1145/3279810.3279850

CCS CONCEPTS

• **Human-centered computing** → *Empirical studies in ubiquitous and mobile computing*;

KEYWORDS

Lifelog, Episodic Memory, Recollection, Experience Sampling, Ubiquitous Computing, Wearables, Cognitive Modeling

ACM Reference Format:
Bernd Dudzik, Joost Broekens, Mark Neerincx, Jeffrey Olenick, Chu-Hsiang Chang, Steve W. J. Kozlowski, and Hayley Hung. 2018. Discovering Digital Representations for Remembered Episodes from Lifelog Data. In *Workshop on Modeling Cognitive Processes from Multimodal Data (MCPMD'18), October 16, 2018, Boulder, CO, USA.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3279810.3279850

1 INTRODUCTION

Experience Sampling Methods (ESMs) refer to a variety of approaches used by researchers for collecting self-reports (e.g. with questionnaires) from individuals about their subjective impressions, thoughts and feelings in the scope of their everyday lives [14]. Some studies have used these methods for the collection of data detailing subjects' experiences during specific situations, e.g. social interactions [12] or instances in which addicts experience craving [17].

Recently, studies have begun to combine this form of data collection with ubiquitous monitoring via wearable sensors, e.g. to investigate long-term team dynamics [11]. Such devices offer additional information about the behaviors displayed by participants, as well as their corresponding context. In combination, these two sources of information hold the potential to provide researchers with a detailed description of how complex social and psychological phenomena emerge and evolve over time [16].

However, an open challenge to unlocking the full potential offered by such a synchronized description is to unpack which sensor readings describe those moments in time that individuals are referring to in their self-reports. This is a difficult task, because of what we will refer to in the following as the *recollection gap*: in contrast to sensor data, the subjective impressions that people are sharing in this way are not referencing objective time periods. Instead, these are grounded in the recollections of their past as specific *Episodes*. These are mental constructs comprising slices of their previous experience. They are primarily defined in terms of their content (i.e. *"what"* they are about) [19], as well as their relative position within the remembering person's overarching life story [4].

While it is possible for people to provide an objective time for the episodes that they remember, it appears difficult for them to do so accurately or consistently (see e.g. [17]). Consequently, time-based information alone is of limited help in bridging this gap. Instead, we need to find a way of situating episodes within an objective timeline, based on those attributes that define them for the person undergoing recollection: elements of the episodic content experienced and associations with their personal history.

Lifelogging is a special approach to ubiquitous monitoring that can contribute towards such a human-centered approach for bridging the recollection gap. Instead of merely organizing data into a timeline, lifelogging provides automatically-generated semantic annotations along-side it. These are meant to approximate an individual's subjective impressions in the situations that he or she encounters while being monitored [9]. For example, a person may be equipped with a wearable camera whose recorded images are then automatically annotated with the labels of places or objects that are visible in them. Because these labels are based on data that was captured from a the subjective point-of-view of the person, they may act as meaningful proxies for the person's actual perceptions. To highlight this connection, we will explicitly refer to annotations created in such a fashion as *Perception Proxies*.

Importantly, these proxies may support anchoring remembered episodes in an objective timeline: the places, people, or objects that an individual experiences as part of an episode, may possess corresponding proxies within their collected lifelog timeline. Consequently, a segment of this timeline that corresponds with content of the recollected episode, may serve as a plausible representation for it. In essence, such *Digital Episode Representations (DERs)* allow an estimate of *when* a given episode may have occurred, and for *how long* it may have lasted.

In this article, we propose a novel approach for bridging the recollection gap by discovering such plausible representations for episodes of interest from lifelog data. In essence, it takes the form of a computational model of the memory processes that have resulted in the recollection of these specific episodes: when provided with a description of a target episode (an indication what was remembered by the person), it emulates the process leading to its recollection by extracting some segment from the lifelog (an indication of what has been experienced) that corresponds with it.

With respect to this, our primary contributions in this paper are the following:

- We present an approach for computationally modeling individuals' memory processes when responding to specific requests for information about their past.
- We give a detailed explanation of a computational model for the specific memory processes displayed by the participants in a longitudinal study, reflecting about social interactions with each other.
- We report on a series of empirical investigations in which we explore the behavior of our model for the recollections in this particular scenario, as well as the representations it discovers.

2 RELATED WORK

Important attributes that distinguish lifelogging from other approaches to the pervasive monitoring of individuals (such as surveillance) include: *1)* a focus on passive and continuous capture of data related to a *single individual* [9], *2)* the collection of data from a subjective point-of-view through wearable devices (e.g. [13]), and *3)* a focus on the automatic annotation of data-traces with labels that describe a person's subjective impressions (e.g. by naming places, objects or persons detected in visual data [6, 10]).

Technical approaches to construct lifelogs have adopted *events* as a basic unit of organization for timelines [9]. Different methods have been devised to provide automatic temporal segmentation of multimodal data streams in such a fashion (e.g. [18]). Similarly, research on lifelogging applications has explored the aggregation of semantic annotations in a timeline to provide relevant descriptions at this event-level [20]. However, the goal of such endeavors is not to discover representations for specific episodes. Rather, they try to create meaningful atomic units to manage and access the large collections of personal data that are being produced by lifelogging appliances [7, 9]. That is, their purpose is to facilitate generic information retrieval tasks. As far as we are aware, no other work in the lifelogging-domain has attempted to create digital representations for episodes in the sense that we describe here.

3 OUR APPROACH

In summary, the approach that we propose for discovering representations for remembered episodes consists of two steps:

(1) Constructing a computational model for the specific memory processes that have lead to the recollection of the target episodes. In particular, this involves the specification of a process for evaluating a segment of lifelog data for its correspondence with these episodes.
(2) Applying this model to lifelogs from the individuals that have remembered these target episodes, in order to identify plausible representations for them.

In *Section 4* we provide a general outline of our computational model for memory processes underlying the recollection of episodes when being asked for information about one's past. In *Section 5* we describe a dataset that was obtained as part of a longitudinal study, and contains information describing recollected episodes in addition to relevant lifelog timelines. It forms the context for an empirical investigation of our approach in *Section 6*. There we give an account of our computational model for participants' memory processes in this particular setting, and explore both its behavior its results when discovering DERs.

4 A COMPUTATIONAL MODEL OF MEMORY RESPONSES

Contemporary psychology generally agrees that access to memories describing personal experiences can take two basic forms: either

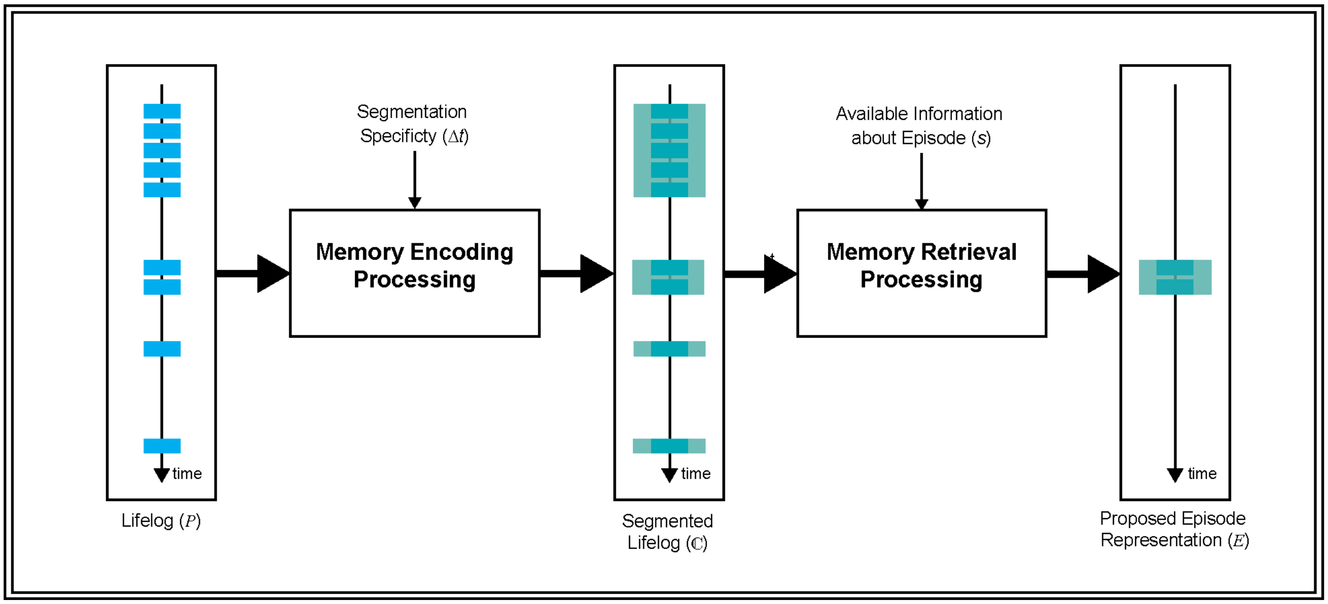


Figure 1: Overview of the proposed model for memory responses for discovering DERs. The *Memory Encoding Processing* splits a lifelog timeline into segments, which are evaluated for their correspondence with information about the episode at the stage of *Memory Retrieval Processing*. The segment with the greatest correspondence is proposed as a candidate for representation.

they emerge on their own, based on associations with cues in one's environment, or one deliberately causes them by searching for information about the past [5]. Requesting someone to provide information about their past, as is done in experience sampling, can be seen as instructing a person to initiate such a deliberate search. In essence, the precise instructions that a person is provided *define* some attributes that an episode needs to fulfill to be considered as relevant for recollection. In the following, we will summarily refer to all the cognitive processes that are undertaken by a person to answer such a request about their past as his or her *memory response*.

In this section, we introduce a computational model of such responses for the purpose of discovering DERs. We will provide a detailed description of the sub-processes constituting it, as well as the representations that it draws on (see *Figure 1* for an overview).

4.1 Memory Encoding Processing

Memory Encoding describes the cognitive process utilized by individuals to parse their continuous experiences into mental representations, which are later accessible as distinct episodes. An important principle in human cognition for integrating experienced stimuli into the same episodes is their consecutive temporal proximity to each other [3, 8]. That is, information that is experienced as occurring relatively close to each other, also tends to be recollected as part of the same episode.

The sub-process of *memory encoding processing* in our model operates according to this specific principle. It's purpose is to emulate the memory encoding that has preceded the recollection of a specific episode in a psychological plausible way.

When provided with a lifelog timeline, it splits it into a collection of non-overlapping segments by grouping temporally close perception proxies together. Each of these segments is then considered to be a potential candidate for representing the outcome of the modeled memory response, i.e. the episode for which a corresponding representation should be discovered.

For this purpose, let $P = \{p_0, p_1, \ldots, p_n\}$ denote a lifelog timeline wherein each element is a timed perception proxy p. A perception proxy itself takes the form of a 3-tuple (t, a, o), were, t is a numerical timestamp that denotes when the entry has been created, a is a label that describes the content that it stands in for (e.g. the name of a specific place or object that was encountered by a person), and o is a unique identifier for the person from whose perspective it was created.

The process of memory encoding then is denoted by the function $enc(P, \Delta t)$. It partitions the contents of a lifelog timeline into a collection of non-overlapping segments $\mathbb{C} = \{C_0, C_1, \ldots, C_n\}$. This segmentation is regulated by the parameter Δt that denotes the amount of time that can pass between two consecutive perception proxies in the timeline P, before they are assigned to a different segment (*segmentation specificity*):

$$\forall C \in \mathbb{C}\left(\forall i\left(|t(p_{i+1}) - t(pi)| < \Delta t \wedge p_i \in C \wedge p_{i+1} \in C\right)\right) \tag{1}$$

4.2 Memory Retrieval Processing

This stage approximates those cognitive processes that have resulted in an individual's willful recollection of a specific episode as part of the modeled memory response. When provided with a collection of candidate segments formed from a given lifelog timeline, it assesses the degree to which each such segment corresponds with information that is available about the episode for which a DERs should be discovered.

To this end, it defines a computational evaluation procedure represented by some function $cor(C, s)$. Here $C \in \mathbb{C}$ is a specific candidate segment under evaluation, while s refers to a collection of available information about an episode that the individual has recollected as part of the modeled memory response. The computational procedure for this evaluation of each segment can take any information into account that is provided by the perception proxies in its timeline. The outcome is a numerical score in the interval $[0, 1]$. A result of 0 describes no correspondence with information describing the episode, while a 1 stands for the greatest possible degree of correspondence.

Given this, the lifelog segment that achieves the highest degree of correspondence is chosen as the most plausible candidate for representation of the episode:

$$E = \underset{C \in \mathbb{C}}{\operatorname{argmax}} \left(cor(C, s) \right) \tag{2}$$

5 THE DATASET

The dataset that we use for an empirical exploration of our approach to discover episode representations was collected as part of a longitudinal study about the dynamics of team-cohesion, and has been utilized in previously published work (e.g [21]). It describes the social interactions of six participants (here coded as *P1* to *P6*) within an isolated environment in the context of a simulated space mission.

For our purposes, two types of records that were collected are particularly relevant: *1)* a range of experience samples in which participants reflect about occurrences of social interactions with each other, and *2)* associated lifelog data from the perspective of each participant. In the following we will describe relevant aspects of these records and how they were collected in more detail. Because one of the participants (coded *P5*) withdrew early from the study for personal reasons, we disregard those records entirely from both our description and modeling activities.

5.1 Experience Samples

Participants were instructed to provide structured reports about the occurrence of social interactions twice-daily at fixed times: once in the morning and once in the evening. Reports could be voluntarily provided at any time through a computer-based questionnaire. This questionnaire instructed participants to recollect and evaluate the most recent social interaction that they had engaged in with other members of the team. The information that they were required to provide about this interaction included the identity of their interaction partners. Moreover, each reported instance could also be annotated with one or more labels specifying the type of interaction it pertained to. Choices that participants were provided with included: *Task Interaction related to Team Goals (T)*, *Task Interaction related to Individual Responsibilities (I)* or *Social Interaction (S)*. Additional evaluations that were requested from them involved judgments of their experiences during the interaction, as well as its perceived effectiveness. Additionally, the time at which participants started and completed the form was automatically recorded by the system.

Table 1 lists the experience samples available for each participants.

Table 1: Experience Samples per Participant.

	P1	P2	P3	P4	P6	Total
N	193	197	151	140	190	871

A detailed look at this collection of reports also exemplifies some of the practical challenges of situating episodes within an objective timeline. While each experience sample possesses a timestamp for when itself was provided by a participant, this does not necessarily allow one to demarcate when the remembered episode itself took place. Especially problematic w.r.t. this is that participants appear to often cross the specified sampling intervals when providing their reports. This can be spotted in *Figure 2*: there is an overproportionally large total share of samples present in the second half of a days. This clearly indicates instances in which multiple reports were provided in a narrow range within the same sampling interval, i.e. in the evening. Because of this, it is no longer possible to just use the timestamps associated with any report to situate the episodes that they refer to even at a coarse level of half a day.

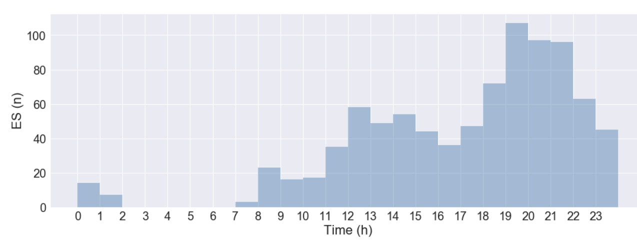


Figure 2: Distribution of the time when participants handed in Experience Samples (hours)

5.2 Lifelog Timeline of Contact Detections

The dataset contains a range of records that have been obtained through pervasive monitoring of participants' behaviors during their daily social interactions throughout the study. These recordings were collected by devices known as *Sociometric Badges*, wearable monitoring platforms that continuously sense their users' relative motion, acoustic ambiance, and the proximity to other badges. For an in-depth description of all the data captured by such a device we refer the reader to [2].

Of relevance for the current study is that badges create a timeline of annotations that uniquely identify any other badge they encounter in close proximity. The devices create this information through a hardware-based process: each device broadcasts a uniquely identifiable infrared signal that can be received by other badges within a reception cone with a 30 degree in a range of about 1.5 meters [1]. Research has demonstrated that this method is reliable at detecting co-location of wearers, but that its ability to do so comprehensively is negatively impacted by barriers and the limited detection range [1]. We will refer to this data for the remainder of this article simply as *Contact Detections*.

Table 2 provides an overview of the total amount of such contact detections that have been registered by the badges of each participant in the study.

Table 2: Contact Detections per Participant.

	P1	P2	P3	P4	P6	Total
N	28194	21840	30018	913	13142	94107

6 EMPIRICAL INVESTIGATIONS

In this section we describe a series of empirical investigations in which we model the memory responses of participants in the previously described study to discover representations for the episodes in our dataset.

As an initial step, we identified properties of the episodes that participants have recollected as part of their memory response which a plausible representation should correspond with. For this we turned towards the data that is available as part of their self-reported descriptions, as well as the instructions that they were provided with. We identified the following two attributes:

- **Presence:** there is a part of participants' self-reports that details exactly which other people were present during the episode that they refer to in their experience sample. This means any plausible DERs should involve references to this group of fellow participants.
- **Recency:** when prompted, participants were explicitly instructed to report the most recent instantiation of what they considered to be a social interaction. Therefor, a plausible representation will need to be situated in temporal proximity to the moment of recall. This moment is documented as part of their self-reported experience samples.

In the following, we first detail how we preprocessed the dataset for usage in our empirical investigations. We then describe the correspondence evaluation function that we modeled for the memory response in this study. Finally, we outline an experiment in which we explore the degree of similarity of the correspondence in the representation that our model is able to discover within- and across-individuals. Finally, we provide an overview of the DERs that our model proposes for the episodes in this dataset.

6.1 Data Preprocessing and Selection

In this section we account for how we preprocessed and selected the elements from the dataset that we deemed relevant for discovering representations that display correspondence with participants' recollections in terms of their *presence* and *recency*.

For this purpose, we use information that was provided by participants as part of their experience samples $S = \{s_0, s_1, \ldots, s_n\}$, and the contact detection-data that was recorded by the sociometric badges of participants. We interpret the latter as a lifelog timeline of perception proxies $P = \{p_0, p_1, \ldots, p_m\}$.

Here, a single experience sample is a record s in the form of a 3-tuple (t, R, o). Were, t, refers to an integer timestamp denoting the time at which a sample was handed in by a participant, while R refers to a set of labels that denote which other participants' were reported as being present in the episode referred to by the experience sample. Finally, o is a label denoting the identity of the participant that is the author of the experience sample.

The available data on contact detections form a lifelog timeline of individual records p that are timed perception proxies for the presence of specific other participants. Each instance of such a proxy is also represented in the form a 3-tuple, (t, a, o). The meaning of t is the time at which the record was created, a is the label of the participant that was detected, and o is the label identifying the participant from who's perspective the proxy was recorded.

6.1.1 Preprocessing Experience Samples. We excluded 22 experience samples from the dataset due to malformed entries, or because they were likely misreports. Additionally, we re-dated some self-reports from within the pool of available samples for participants as part of the preprocessing for our experiment. We modified all reports that were handed in before 3am in the morning and for which no available lifelog data exists for this period from within the same day. In these cases we assumed that a sampling interval had been skipped by participants, i.e. that they had reported an episode from the day before. To more accurately reflect participants' recollection behavior, we associated such samples with the previous day (11:59:59pm).

6.1.2 Preprocessing Lifelog Data. The perception proxies contained in the lifelog timelines of the study are not mutual. This means that there exist instances where one participant's records indicate contact with another person, without that person's sensor producing a matching entry in their own lifelog timeline. However, we assume that co-location at such close range as is being registered by the wearable sensor result in mutual perceptions between participants (i.e. *"If I see you, then you see me as well"*). To reflect this, we mirrored entries across participants' timelines and combined these mirrored versions with the original lifelogs into an extended dataset P^+. It fulfills the following constraints:

$$\forall p_x \in P^+ \forall p_y \in P^+ \Big(p_x \neq p_y \wedge t(p_x) = t(p_y) \wedge o(p_x) = a(p_y) \wedge a(p_x) = o(p_y) \Big) \quad (3)$$

6.1.3 Alignment and Selection. Finally, we partitioned our lifelog dataset P^+ into individual segments, each spanning a period of time in which representations for a specific experience sample should be discovered. This means from the beginning of the same day on which the episode has occurred, up to the moment it was reported. That is, all parts resulting from this partitioning $\{P_0^+, P_1^+, \ldots, P_n^+\}$ fulfill the following constraints:

$$\forall i \Big(\forall p \in P_i^+ \Big(t(p) < t(s_i) \wedge day(p) = day(s_i) \wedge o(p) = o(s_i) \Big) \Big) \quad (4)$$

The result is an aligned dataset that contains parings of participants' experience samples with relevant segments from within their lifelog timelines (P_i^+, s_i).

From the total amount of 861 such pairings, not all did meet our requirements. We removed an additional 220 such pairings, because there was no lifelog data present for the relevant period of time.

Furthermore, we had to remove a set of 70 samples for which there was no overlap between the people that were present in a participant's description of the episode that he/she recollected for the experience sample and the associated lifelog data. Table 3 provides an overview of the remaining data pairings that we used for our experiments split by participants.

Table 3: Final number of data pairs (P_i^+, s_i) selected for usage in our experiments.

	P1	P2	P3	P4	P6	Total
N	152	146	138	24	112	572

6.2 Modeling Participants' Memory Retrieval Processing

In this section we describe our computational approach for assessing the degree with which a lifelog segment displays correspondence with the available information about an episode. As mentioned above, we identified two attributes of the episodes in this scenario that representations will need to meet: the presence of specific other participants, and recency w.r.t. the moment of recall.

To assess the degree to which a lifelog segment corresponds with these properties, we constructed the following evaluation function *cor*:

$$cor(C, P, s) = pres(C, P, s) * rec(C, P, s) \tag{5}$$

where C is a given candidate segment of a participant's lifelog P, s is relevant information about the recollected episode. The total evaluation of a candidate C consists of two partial functions, each of which assesses the degree to which one of the correspondence requirements is met. In our model, plausible representations need to possess both attributes jointly for achieving maximum correspondence.

6.2.1 Presence Evaluation. For assessing the correspondence between a lifelog segment in terms of the people that were reported as present by a participant, we compared the degree to which the labels of the perception proxies it contains match their description in the following way:

$$pres(C, P, s) = \frac{sim(C, s)}{sim_{max}(P, s)} \tag{6}$$

In this function, $sim(C, s)$ is the *Jaccard Similarity* between the set of all annotations describing the presence of participants in the lifelog segment C and the set of labels that denote who was present in the associated self-report s. We normalized this measure over a value generated via the operation $sim_{max}(P, s)$. It denotes the maximum possible overlap between the annotations contained in the lifelog from which the segment under investigation was created$P \supseteq C$ and the self-report s. The reason for this procedure is that there are cases in which not all individuals that were reported as present were also detected within the relevant lifelog. This may be a result of the rather short detection range of the sociometric badges, causing participants not to be registered, even though they are perceived as present. Together, this function provides a relative measure of a segment's correspondence w.r.t. the presence of other participants in the range from $[0, 1]$. A 0 denotes a total discrepancy between the two accounts, while a 1 forms the best match possible for a representation created from this particular lifelog timeline.

6.2.2 Recency Evaluation. Next, we devised an evaluation function to assess the degree to which a lifelog segment C under evaluation displays recency w.r.t. the moment at which the memory response took place, as indicated by the timestamp in the associated self-report s:

$$rec(C, P, s) = 1 - \Big((t_{rel}(s, P) - t_{rel}(C, P)\Big) \tag{7}$$

In essence, this function provides a measure between the time when a self-report was provided, and the beginning of the lifelog segment under evaluation (i.e. the timestamp of the first perception proxy). Importantly, these moments transformed to their relative position within the timeline of the lifelog from which the segment was created $P \supseteq C$. This is achieved by normalizing both objective timestamps over the duration that is covered by the lifelog timeline, an operation that is denoted by $time_{rel}$. The resulting overall measure for recency for any given segment under evaluation falls within the interval $[0, 1]$, where a 0 denotes a maximally distant segment (i.e. it is located at the furthers point away in the timeline of the lifelog), while a 1 is a maximally recent one (it is the closest point in the relative timeline of the lifelog).

6.3 Exploration of Similarity in Representation Discovery

An implicit assumption of our model for the memory responses in this study is that they are highly similar to each other. That is: prompting individuals to remember experiences in their past using the same prompt is assumed to result in a very similar form of recollection for each instance, independently of who is confronted with it, or when that is. Arguably, the existence of such a shared memory response is an essential property for experience sampling. Without it, these methods would not be able to provide comparable information from different participants in a study and at different moments in time.

In this section we explore whether our model would display a behavior that reflects this property when discovering correspondent representations for episodes in this study.

To gain insights into this, we conducted three experiments using our preprocessed dataset in different cross-validation schemes. These allowed us, to study the degree to which a model that was trained to reflect the memory responses of some subset of our data, would vary in the correspondence that it produces when being applied to unseen instances.

6.3.1 Experimental Setup. For the purpose of this exploration we devised the following three cross-validation schemes:

- **WithinCV:** For each participant we partitioned all available pairings of (P_i^+, s_i) into five segments. Each segment was populated via random sampling without replacement. We

Table 4: Average Results for all Experiments

Type	N	AvgCor±SD ($\Delta Train$)	AvgPres±SD ($\Delta Train$)	AvgRec ±SD ($\Delta Train$)
WithinCV	25	$.60 \pm .09\ (-.02)$	$.73 \pm .07\ (-.01)$	$.85 \pm .08\ (-.01)$
StratCV	5	$.59 \pm .03\ (-.01)$	$.72 \pm .03\ (-.01)$	$.85 \pm .04\ (< .00)$
LopoCV	5	$.58 \pm .02\ (-.02)$	$.73 \pm .07\ (+.01)$	$.83 \pm .07\ (-.02)$

used this division in a 5-Fold Cross-Validation procedure for training and testing of a model for each individual.

- **StratCV:** This experiment involves training and testing with a 5-Fold Cross-Validation procedure. Each partition is populated by randomly selecting pairings (P_i^+, s_i) without replacement. The amount of pairings that are selected from each participant's data to populate a segment is proportional to their share in the overall amount.
- **LopoCV:** In this experiment, we split all available pairings (P_i^+, s_i) into 5 segments. Each consists of all the data associated with a specific individual in the study. Training is then undertaken in a Leave-One-Participant-Out fashion. That means, we first train our model on data of 4 participants, and then apply it to the held-out data from the remaining individual.

The goal of the WithinCV-procedure was to gain insights into the similarity of the correspondence-scores produced by our model when trained and tested based on instances belonging to the same individual. In contrast, both StratCV and LopoCV provide insights into the consistency of the correspondence displayed by our model for instances belonging to different participants in the study.

In order to reflect the memory responses underlying the episodes described in the study, we train our model to learn a parameter Δt that maximizes the average correspondence of proposed representations over all available pairs of data (P_i^+, s_i) that were assigned to a particular segment of the training-data:

$$\underset{\Delta t}{\text{argmax}}\ \frac{1}{n} \sum^{i=n} \left(\max_{C \in enc(P_i^+, \Delta t)} cor(C, P_i^+, s_i) \right) \tag{8}$$

Since in our scenario the timeline spanned by lifelogs consists of only a single day, we optimized correspondence during training with a sweep Δt over the interval $[0, 20000]$ (seconds). This means, that two consecutive perception proxies in the lifelog cannot not be farther apart than $5\frac{1}{2}$ hours from each other to be counted towards the same segment. In situations where multiple optimal solutions for Δt were discovered in a training phase, we selected the one with the smallest value. This corresponds with a preference for models with a more specific segmentation over broader ones.

6.3.2 Results. The information in *Table 4* represents the average results that were achieved in these experiments (i.e. averaged over all folds). The optimized average correspondence for DERs achieved by our model varied only minimally between the testing and training phases ($\Delta Train$). This is the case independently of whether it was trained to reflect memory responses within a single participant, or when spanning data from different persons. Moreover, both the recency and presence components that comprise these correspondence scores display such a similarity. We see in this behavior a property that one would expect in an experience sampling scenario, i.e. a substantial degree of similarity across all instances of the memory responses. This adds further plausibility to the representations that are discovered by our model for the memory response in this scenario.

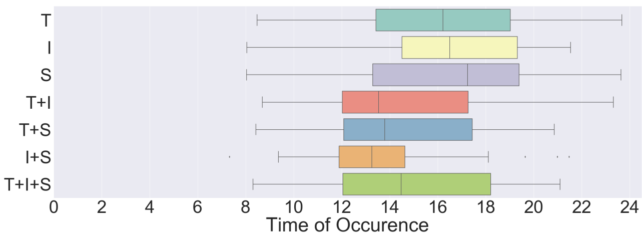
6.4 Exploration of Discovered Episode Representations

In this section we explore the DERs that were discovered by our model for the recollections of participants in this study when trained in a person-independent fashion on all available pairings (P_i^+, s_i). The discussions in this section are not intended to provide a thorough analysis of participants' social interactions. Instead, they form a demonstration of the insights that possession of DERs could provide to support researchers that undertake such an endeavor.

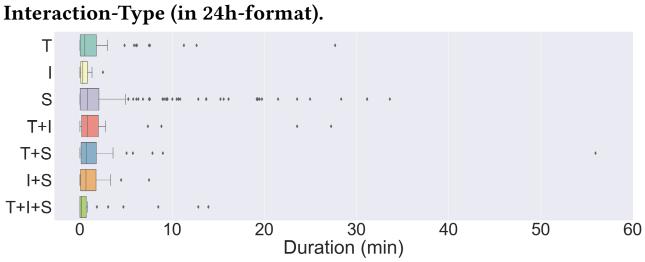
6.4.1 Time of Occurrence. Most of the discovered DERs are located in the afternoon ($M = 15.753.83$, $SD = 3.83$, $N = 571$), but they cover the entire waking day period of participants. *Figure 3a* describes the distribution of where episode representations are situated, sorted according to how participants labeled the interactions during them. A potential pattern that can be spotted when looking at this distribution relates to representations for episodes which revolve around a mixture of individual responsibilities and socializing (i.e. I+S-type interactions). These are generally situated at midday ($M = 13.65$, $SD = 3.22$, $N = 30$). This is not the case for representations of episodes that are perceived as either being purely social (S-type interactions, $M = 16.23, SD = 3.86, N = 343$), or to entirely revolve around work (I-type interactions, $M = 16.18$, $SD = 4.01$, $N = 11$). Both of these tend to be situated rather later in the day. Together, this could indicate that the activities spanned by I+S interactions describe meetings where individual tasks were discussed among team-members over shared meals around lunchtime.

6.4.2 Duration. The average duration of the DERs discovered by our approach was $M = 2.25$ minutes ($SD = 5.12, N = 571$). *Figure 3b* describes their distribution according to the associated interaction-type. The discovered DERs for purely social interactions (S-type) tend to have the longest average duration ($M = 2.42$, $SD = 5.05, N = 343$). On the other hand those that were characterized as revolving around individual responsibilities take up the shortest average amount of time ($M = 0.62$, $SD = 0.74, N = 11$). This could be a result of the strongly task-oriented nature that participants ascribe to these interactions, reflecting short and efficient discussions.

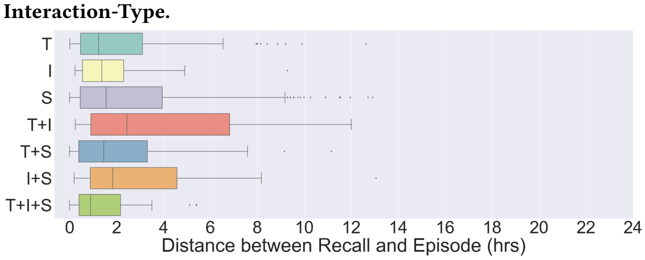
6.4.3 Distance to Recall. Another interesting aspect of the discovered DERs is their relative distance to the point in time at which participants provided a corresponding self-report (see *Figure 3c* for



(a) Distribution of the Time of Occurrences for DERs, sorted by Interaction-Type (in 24h-format).



(b) Distribution of the Duration of discovered DERs(Minutes) by Interaction-Type.



(c) Distribution of the delay between the discovered DERs and the moment of participants' recollection (in Hrs).

Figure 3: Discovered Representations by Interaction Type. Labels refer to T: Task Interaction related to Team Goals, I: Task Interaction related to Individual Responsibilities, S: Social Interaction. Labels combined with a '+' represent interactions that were labeled as mixed by participants.

the distribution according to Interaction-Type). On average, representations are situated around three hours before a participant's self-report ($M = 2.71, SD = 2.91, N = 571$). Such information could, for example, be helpful in identifying an opportune structure for requesting self-reports in a study.

7 SUMMARY AND CONCLUSION

The combination of ubiquitous monitoring and self-reported reflections into a synchronized timeline has potential for increasing our understanding of how behavior emerges and unfolds in the scope of everyday lives. We have argued that one principal challenge that needs to be addressed to make progress towards providing such a synchronized description, is to organize data in a fashion that is analogous to how individuals experience their personal past in recollection.

In this article, we have suggested that lifelogs form a meaningful source for representations to anchor remembered episodes within an objective timeline. To this end, we have described an approach for discovering candidates for such representations by computationally modeling the memory responses underlying their recollection. We have applied this approach to a dataset describing recollections of participants in a longitudinal study, and have argued that this has resulted in plausible representations for them. Our brief exploration of these representations has hinted at some of the insights that might be gained about individuals' social interactions through their study.

Undertaking our empirical investigation has revealed several opportunities for further exploration. First, while we consider the discovered representations in our scenario as plausible, we did not demonstrate that they are also *accurate*. That is, we have not provided empirical evidence for the degree to which their estimated position in a timeline corresponds with the period referred to by participants when providing a self-report. Future research might explore ways of conducting such evaluations, as well as the collection of relevant data for it. Second, while annotations in a lifelog timeline have the potential to indicate information that *could have* been perceived by monitored individuals, they are not guaranteed to reflect what actually *was* perceived by them. This is primarily caused by their inability to mirror human attentional processes when creating annotations. In our opinion, this forms a general challenge for lifelogging as a research field. A starting point for addressing it may be found in existing research that explores the computational modeling of human attentional processes [15].

In summary, we see our approach as a contribution towards enabling ubiquitous computing applications to create synchronized descriptions that reflect how people experience their daily lives, as well as how they behave in them. In our opinion, the information provided by lifelogs forms a valuable resource for bridging the gap between remembered experience and objectively collected data, and its potential in this respect should be the target of further research.

ACKNOWLEDGMENTS

Data used in the preparation of this manuscript were provided by S. W. J. Kozlowski, S. Biswas, and C.- H. Chang. Their research, which generated the data, was supported by the National Aeronautics and Space Administration (NASA; NNX13AM77G, S.W.J. Kozlowski, Principal Investigator). Any opinions, findings, conclusions and recommendations expressed are those of the authors and do not necessarily reflect the views of NASA.

REFERENCES

[1] Daniel Chaffin, Ralph Heidl, John R Hollenbeck, Michael Howe, Andrew Yu, Clay Voorhees, and Roger Calantone. 2017. The Promise and Perils of Wearable Sensors in Organizational Research. *Organizational Research Methods* 20, 1 (jan 2017), 3–31. https://doi.org/10.1177/1094428115617004

[2] Tanzeem Choudhury and Alex (Sandy) Pentland. 2002. The sociometer: A wearable device for understanding human networks. *CSCW'02 Workshop: Ad hoc Communications and Collaboration in Ubiquitous Computing Environments, New Orleans, Louisiana, USA* (2002). https://doi.org/10.1.1.57.9810

[3] David Clewett and Lila Davachi. 2017. The ebb and flow of experience determines the temporal structure of memory. *Current Opinion in Behavioral Sciences* 17 (oct 2017), 186–193. https://doi.org/10.1016/j.cobeha.2017.08.013

[4] Martin A. Conway. 2009. Episodic memories. *Neuropsychologia* 47, 11 (sep 2009), 2305–2313. https://doi.org/10.1016/j.neuropsychologia.2009.02.003

[5] Martin A Conway and Christopher W Pleydell-Pearce. 2000. The construction of autobiographical memories in the self-memory system. *Psychological Review* 107, 2 (apr 2000), 261–288. https://doi.org/10.1037/0033-295X.107.2.261

[6] Aiden R. Doherty and Alan F. Smeaton. 2008. Combining face detection and novelty to identify important events in a visual lifelog. *Proceedings - 8th IEEE*

International Conference on Computer and Information Technology Workshops, CIT Workshops 2008 (2008), 348–353. https://doi.org/10.1109/CIT.2008.Workshops.31
[7] Aiden R Doherty, Alan F Smeaton, Keansub Lee, and Daniel P W Ellis. 2007. Multimodal segmentation of lifelog data. *RIAO 2007 LargeScale Semantic Access to Content Text Image Video and Sound* 2006 (2007), 21–38.
[8] Simon Farrell. 2012. Temporal clustering and sequencing in short-term memory and episodic memory. *Psychological Review* 119, 2 (apr 2012), 223–271. https://doi.org/10.1037/a0027371
[9] Cathal Gurrin, Alan F. Smeaton, and Aiden R. Doherty. 2014. LifeLogging: Personal Big Data. *Foundations and Trends® in Information Retrieval* 8, 1 (2014), 1–125. https://doi.org/10.1561/1500000033
[10] Cathal Gurrin, Alan F. Smeaton, Zhengwei Qiu, and Aiden Doherty. 2013. Exploring the technical challenges of large-scale lifelogging. In *Proceedings of the 4th International SenseCam & Pervasive Imaging Conference on - SenseCam '13*. ACM Press, New York, New York, USA, 68–75. https://doi.org/10.1145/2526667.2526678
[11] Steve W J Kozlowski, Georgia T Chao, Chu Hsiang Chang, and Rosemarie Fernandez. 2015. *Big Data at Work*. Routledge. 272–309 pages. https://doi.org/10.4324/9781315780504
[12] Jean-Philippe P Laurenceau, Lisa Feldman Barrett, and Paula R. Pietromonaco. 1998. Intimacy as an interpersonal process: the importance of self-disclosure, partner disclosure, and perceived partner responsiveness in interpersonal exchanges. *Journal of personality and social psychology* 74, 5 (may 1998), 1238–51.
[13] 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*, Vol. 344. ACM Press, New York, New York, USA, 44. https://doi.org/10.1145/1409635.1409643
[14] Christie Napa Scollon, Chu-Kim Prieto, and Ed Diener. 2009. Experience Sampling: Promises and Pitfalls, Strength and Weaknesses. In *Assessing Well-Being: The Collected Works of Ed Diener*, Ed Diener (Ed.). Vol. 4. Springer, Dordrecht, 157–180. https://doi.org/10.1007/978-90-481-2354-4_8
[15] Claudia Roda and Julie Thomas. 2006. Attention aware systems: Theories, applications, and research agenda. *Computers in Human Behavior* 22, 4 (jul 2006), 557–587. https://doi.org/10.1016/j.chb.2005.12.005
[16] Eduardo Salas, Rebecca Grossman, Ashley M. Hughes, and Chris W. Coultas. 2015. Measuring Team Cohesion: Observations from the Science. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 57, 3 (may 2015), 365–374. https://doi.org/10.1177/0018720815578267
[17] Saul Shiffman, Michael Hufford, Mary Hickcox, Jean A Paty, Maryann Gnys, and Jon D Kassel. 1997. Remember that? A comparison of real-time versus retrospective recall of smoking lapses. *Journal of Consulting and Clinical Psychology* 65, 2 (apr 1997), 292–300. https://doi.org/10.1037/0022-006X.65.2.292.a
[18] Estefania Talavera, Mariella Dimiccoli, Marc Bolaños, Maedeh Aghaei, and Petia Radeva. 2015. R-Clustering for Egocentric Video Segmentation. In *Pattern Recognition and Image Analysis*, Paredes R., Cardoso J., and Pardo X (Eds.). Springer International Publishing, 327–336. https://doi.org/10.1007/978-3-319-19390-8_37
[19] Endel Tulving. 2002. Episodic Memory: From Mind to Brain. *Annual Review of Psychology* 53, 1 (feb 2002), 1–25. https://doi.org/10.1146/annurev.psych.53.100901.135114
[20] Peng Wang and Alan F. Smeaton. 2011. Aggregating semantic concepts for event representation in lifelogging. In *Proceedings of the International Workshop on Semantic Web Information Management - SWIM '11*. ACM Press, New York, New York, USA, 1–6. https://doi.org/10.1145/1999299.1999307
[21] Yanxia Zhang, Jeffrey Olenick, Chu-Hsiang Chang, Steve W J Kozlowski, and Hayley Hung. 2018. The I in Team. In *Proceedings of the 2018 Conference on Human Information Interaction&Retrieval - IUI '18 (IUI '18)*. ACM Press, New York, New York, USA, 421–426. https://doi.org/10.1145/3172944.3172997