Measuring the Effect of Cued Recall on Work Meetings

Agon Bexheti
Università della Svizzera italiana
Via Giuseppe Buffi 13,
6900 Lugano, Switzerland
agon.bexheti@usi.ch

Marc Langheinrich
Università della Svizzera italiana
Via Giuseppe Buffi 13,
6900 Lugano, Switzerland
marc.langheinrich@usi.ch

Evangelos Niforatos
Università della Svizzera italiana
Via Giuseppe Buffi 13,
6900 Lugano, Switzerland
evangelos.niforatos@usi.ch

Fabio Crestani
Università della Svizzera italiana
Via Giuseppe Buffi 13,
6900 Lugano, Switzerland
fabio.crestani@usi.ch

Seyed Ali Bahrainian
Università della Svizzera italiana
Via Giuseppe Buffi 13,
6900 Lugano, Switzerland
seyed.ali.bahreinian@usi.ch

UbiComp/ISWC '16 Adjunct, September 12-16, 2016, Heidelberg, Germany.

 ISBN 978-1-4503-4462-3/16/09...$15.00
DOI: http://dx.doi.org/10.1145/2968219.2968563

Abstract
Recent technological improvements allow us to capture an increasing share of our everyday experiences, e.g. holidays, shopping routines, or sports activities, and store them in a digital format. An interesting avenue to explore in this context is how reviewing such captured content can improve one's memories of the original events. In this position paper, we describe a planned experiment to investigate the impact of such captured recordings (and their subsequent review) on supporting work meetings. We provide the planned study procedure, explain the envisioned apparatus and metric, and describe the technology used to support the review activity.

Author Keywords
Human Memory Augmentation; Autobiographical memories; Memory cues; Life logging; Meeting Capture; Meeting Summarization; Topic Modeling.

ACM Classification Keywords
H.5.m [Information interfaces and presentation]: Miscellaneous

Background
Pervasive computing allows people to better manage – and eventually improve – their lifestyle by measuring ("quantifying") many aspects of their daily activities e.g., from steps

taken, to calories burned, to water intake, etc. An interesting application of such captured data is to use it for improving one's memory of past events. The idea behind such a *pervasive memory augmentation system* [3] is a four-step process (see Figure 1): using captured data of a user's everyday experience (step 1), we can generate a set of *memory cues*[1] (step 2) that can be played back to the user in an ambient fashion, e.g., as a computer screensaver or a smartwatch card (step 3), in order to ultimately improve one's own memory of these events.

In order to verify this general idea, we are interested in creating experiments that allow us to measurably improve one's memory of everyday events. One interesting application area are *work meetings*, as people are generally very much interested in remembering them better, yet are often unwilling to take extensive notes or review detailed minutes before subsequent meetings. We thus wanted to set up an experiment that lets us measure how much such a system could improve one's recollection of past meetings.

Current approaches to meeting capture usually require someone to manually take notes (minutes). However, not everybody is prepared to do this during a meeting, as it also affects the flow of the meeting itself. Simply recording a meeting, e.g., using a voice recorder, alleviates participants from note taking, yet few participants are willing to listen to (or watch) the entire recording again. Obviously, such recordings (or any transcript derived from them) do have a value when it comes to refreshing one's memory of a specific targeted point in the meeting, yet they are difficult to use for helping one to remember a meeting as a whole. The best option is probably to have a designated note taker who does not participate in the discussion but who is able

[1] *A memory cue* can be any item e.g., a photograph, a flash card, an audio snippet or even a smell, that reminds you of a past experience.

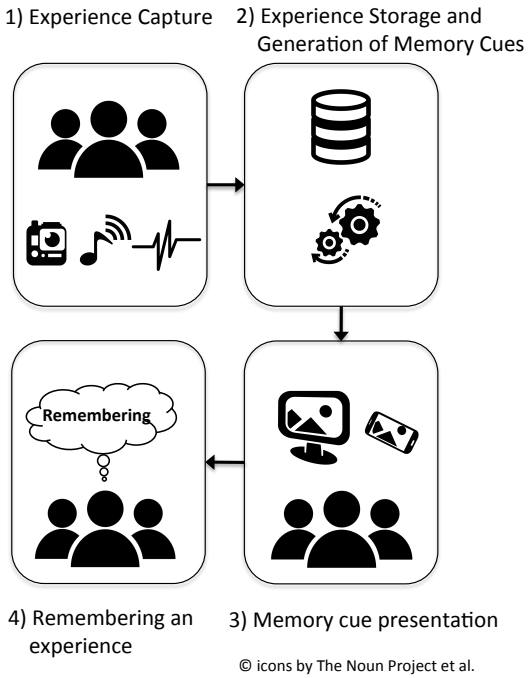

Figure 1: Memory augmentation process.

to summarize the discussion and extract noteworthy issues out of it, in order to keep the resulting record brief and to the point, yet this may not only be too expensive for most meetings but may still require active study of the meeting extract ahead of time. We believe that a memory augmentation system can play a significant role here without having the burden of manually taking notes or spending long time watching meeting recordings. The vision is that one simply reviews few "memory cues" (e.g., in the form of slides) prior to the next meeting (or anytime in between) in order to refresh her memory.

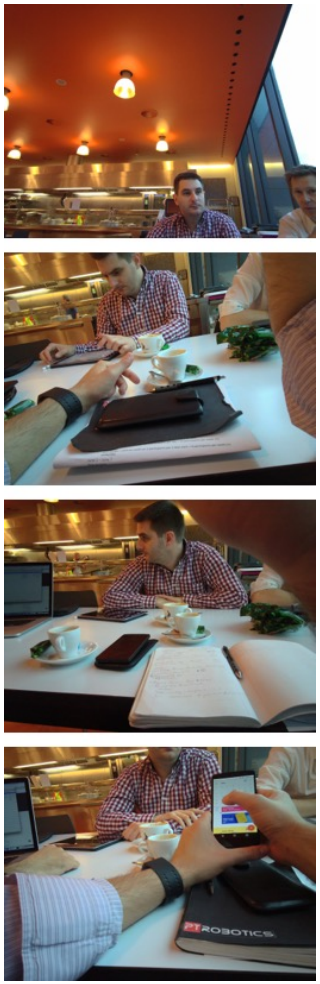

Figure 2: First-person view images as captured by the Narrative Clip.

There is a great body of research focusing on different ways for improving efficiency of meetings. However only few of them include participants' memory augmentation in their research scope. Jaimes et al. [5] present a technique that helps meeting attendees retrieve segments from meeting video recordings based on the use of some high-level search attributes (which they call "memory cues"). They support a finite set of these attributes (e.g., time when meeting happened, meeting room layout, meeting participants, use of equipment in a meeting), which they identified in a previous user study as most memorable to meeting participants. Note that the authors' concept of a memory cue as "something that you remember" is very different from our use of memory cues as "something that reminds you". MemTable [4] is an interactive tabletop system for capturing and reviewing meeting group discussions. It features several input modalities (e.g., simultaneous drawing, text entry, audio recording, image capture) to support heterogeneous collaboration styles. Users can then review such input from previous meetings (which are projected on a tabletop's surface) by searching for particular items using text entries or browsing contents with a timeline.
Both these systems can help participants to better remember previous meetings by allowing them to easily search (or browse) through a rich set of captured contents. However, instead of creating a look-up-like memory aid, we try to understand whether we can use captured data of meetings to *improve* human memory, so that details about a previous meeting can be recalled without an external tool.

RECALL Technology

Our work is part of the EU-funded research project RECALL[2], which aims to re-define and re-think the notion of memory augmentation by combining today's improvements in pervasive computing together with contemporary psychology theories. In a nutshell, we try to understand the effect of reviewing captured and carefully selected memory cues of a previous experience on the ability to recall memories corresponding to that experience (a process known as *cued recall*). For this reason, we build memory augmentation prototypes customized for different domains, e.g. reinforcing learning of new skills, behavior change, supporting failing memories, etc., and try evaluate their effect in real-life scenarios.

RECALL for Work Meetings

In this work we focus on *work meetings* scenario, i.e., employing a memory augmentation system in the context of work meetings and measuring the effect of cued recall on participants' ability to remember past meetings. To study this, we plan to capture a series of weekly meetings of small groups. Each meeting will be captured using state-of-the-art capture technology (e.g., video cameras, wearable cameras, heart rate sensors), which is then processed by our RECALL system to yield a number of visual "memory cues" for each participant. A typical memory cue will be a picture from the wearable camera (see Figure 2) together with a few keywords extracted from the discussion at that time. By briefly reviewing these cues, we expect study participants to have significantly better memories of their past meetings.

Creating Memory Cues for Meetings

The memory cues we envision for the experiment are a combination of meaningful pictures and several keywords indicative of the topic discussed at the time. We automatically create such cues as follows.

At the outset, we use an audio recording of the meeting to transcribe detailed meeting minutes. While this can in principle be fully automated, current speech recognition

[2] http://www.recall-fet.eu

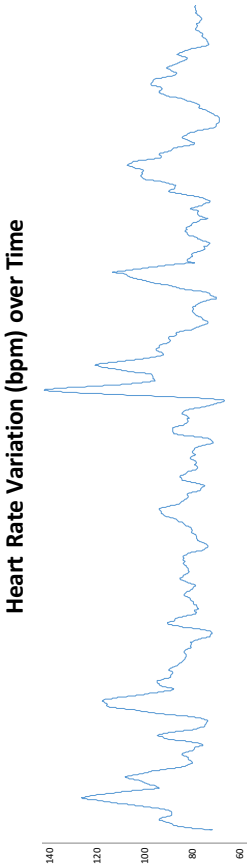

Figure 3: Sample of the Heart Rate Variation (HRV) signal captured from Empatica E4.

technology does not yet offer the quality level required for our cues. We hence resort to manual[3] transcription. The result is a time-stamped transcript of the meeting's discussion with speaker identification (assigning participants' pseudonym IDs). We then process the transcription using a state-of-the-art topic modeling method, Latent Dirichlet Allocation [2], to extract the salient themes of the conversation as a textual summary. Topic modeling is one method for exploring and indexing unstructured datasets of digitized text documents, and it has been used in many different applications such as summarization and trend analysis [1], as well as information retrieval [7]. Topic models are hierarchical Bayesian models of discrete data [6], where each topic is a set of words, drawn from a fixed vocabulary, which together represent a high level concept. LDA uses co-occurrence of words in order to generate topics each representing a high level concept, e.g., participants talking about a research paper, or a discussion on upcoming student projects, which consist of set of words drawn from the entire vocabulary present in the conversation. Depending on the length of the meeting, our algorithm extracts 4-10 such "topics", each being characterized by a set of keywords (e.g., 10) and their relative frequency in this topic.

To create more stimulating cues, we enrich each topic with a first-person view picture captured by the participants' wearable camera. Images are selected based on visual features (sharpness, number of faces in the image, uniqueness within period) and the biophysical data captured through a wristband (see next section for a detailed description of the capture devices used). We then combine both topic words and pictures into a simple slideshow – one slide for each identified topic in the meeting. The choice of first-person view pictures over third-person view pictures was driven by the fact that a wearable camera better captures an event as seen from the wearer's vantage point, thus increasing their power as a personal memory cue. While head-worn cameras would provide an even better approximation of a user's visual memory, we previously found that head-mounted cameras were less stable than wearable cameras (thus creating more blurred pictures) [8]. Moreover, head-mounted cameras are socially more awkward to wear and hence are more difficult to introduce as a non-invasive tool to augment a meeting.

[3] We use a professional transcription service in our experiment.

Experimental Apparatus

Our meeting capture apparatus consists of *Narrative Clip 2* wearable cameras[4] that capture first-person view photographs (one for each attendee), *Empatica E4* wristbands[5] (one per participant) that can record participants' bio-physiological responses, and a single *GoPro Hero 4* action camera[6] to both capture audio and to record the event from a third-person view angle. The apparatus as a whole is highly portable and it can be easily deployed to fit any meeting room configuration.

Figure 4: Left: Narrative Clip 2; Right: Empatica E4.

[4] http://getnarrative.com/
[5] https://www.empatica.com/e4-wristband
[6] https://gopro.com/

The Narrative Clip 2 (see Figure 4 left) is a tiny modern life-logging camera (weights less than 20g) with an 8MP camera that can record both high-resolution photographs (with a frequency between 10s and 30s) and short video snippets (one can choose between lengths of 10s, 20s or 30s) and a WiFi module for easily grabbing its captured content. Its battery allows up to 30h of data capturing.

Figure 5: GoPro Hero 4 with an external directional microphone and a WiFi remote controller.

The Empatica E4 (see Figure 4 right) is a watch sized wristband that can capture users' bio-physiological responses with near to clinical quality resolution. It features a continuous heart rate variation (HRV) sensor, an electrodermal activity (EDA) sensor, a thermostat and a 3D accelerometer. It can record up to 36h of data both in terms of battery life and memory capacity.

The GoPro Hero4 (see Figure 5) has wide-angle camera lenses for capturing up to 4K resolution videos and 12MP photographs with 4000x3000 pixels. We included an external directional microphone to support capturing understandable audio contents of meetings taking place in noisy environments. We also use the optional GoPro WiFi remote control, in order to allow participants to exclude potentially sensitive discussion from being captured by stopping and re-starting capture. While we do not envision the meetings to discuss any sensitive content (we mostly target meetings between student summer interns and their supervisors) we believe this is an important part of making participants feel comfortable signing up. Note that the GoPro is predominantly used as an audio recorder – the video recording is mainly to help with disambiguating speakers during the transcription phase.

Envisioned Study Design

We plan to run a within-subjects design (all participants will be engaged in all conditions) with two conditions, each repeated twice.

Condition A (control): Just before the start of a new subsequent meeting, each meeting participant is, individually, asked to briefly think for 2 minutes about their last meeting. They are then interviewed by a researcher following an adapted version of the *'cognitive interview'*[7] protocol in order to capture their memory of the last meeting. In a cognitive interview, the interviewer simply repeatedly prompts participants to recall more and more details ("And then what happened?").

Condition B (recall): In contrast to the control condition, participants are asked to review a set of memory cues (in the form of a slide deck) for 2 minutes, before conducting the cognitive interview.

Each set of meeting participants will be exposed to conditions A and B for two times each. The lines below provide

[7] Cognitive interview is an effective method usually used by crime investigators to interview eyewitnesses by trying to minimize misinterpretation and uncertainty that is manifested when recalling a previous event.

an overview of the six possible orders of conditions A and B. Each new set of participants will use a different line (i.e., different order), in order to counterbalance conditions. The lines below indicate a meeting capture with *'m'*, the experimental conditions with either *A* or *B*, and and the cognitive interview with *'e'* (for "evaluation"):

- (m)-(Aem)-(Aem)-(Bem)-(Be)
- (m)-(Bem)-(Bem)-(Aem)-(Ae)
- (m)-(Aem)-(Bem)-(Aem)-(Be)
- (m)-(Bem)-(Aem)-(Bem)-(Ae)
- (m)-(Aem)-(Bem)-(Bem)-(Ae)
- (m)-(Bem)-(Aem)-(Aem)-(Be)

We are currently in the process of recruiting a number of small groups (typically 2-3 participants per group) that consent to having us capture their weekly meetings and examining the effect of RECALL on their memory. Each group will be engaged in the study for five consecutive weeks so that we can capture four weekly meetings of theirs. The fact that the experimental prototype is portable and can be easily setup allows us to be more flexible and bring the equipment to groups' premises for each meeting.

Measuring Recall

The set of memory cues representing the participants' last meeting will be delivered to participants in the recall condition as a slide-show (comprising no more than six slides) and presented to them using one of the researcher's laptop device. Participants have 2 minutes to review them. When in the control condition, we simply ask participants to think about the last meeting for an equal amount of time. The choice of 2 minutes is arbitrary, though in our view represents a realistic time estimate for the amount of time spent preparing a subsequent meeting. It is only important that both conditions use the same amount of time.

In both conditions, we will then subsequently ask participants to recall their last meeting. Instead of simply asking them to write down all that they remember, we decided to use a memory elicitation technique known as the *cognitive interview* to both minimize the extra effort our setup imposes on our participants' usual meeting routine, and to elicit more detailed descriptions. A spoken interview not only makes it less of an effort to share one's memories, it also allows us to control for the different styles that people have when recalling past events: if a person is in generally not very detailed storyteller, the interviewer can simply ask for more detail.

In an ideal case, the recall intervention about the last meeting should be carried in an ambient fashion (e.g., memory cues of the last meeting displayed as a screensaver) anytime before the subsequent meeting. However, it is not easy to control for such an intervention (Did people really look at those ambient displays? For how long? How often?), making it difficult to reliably measure its effect. We hence decided to use a single, mandatory intervention in the form of a 2-minute slide review. When should this intervention occur? Most users will prepare for a subsequent meeting few hours, if not minutes, before their next meeting starts. We therefor decided to do the recall intervention (or the self-reflection in the case of control condition) shortly before the next meeting. This should reflect a more realistic use case.

At the end of a meeting, participants are also asked to fill out a short (1-page) questionnaire regarding their subjective impressions of the meeting (e.g., how exciting or boring the meeting felt), and their perception on how useful the recall exercise before the meeting (either with or without the slides) had been for conducting the meeting. Note that

both the recall exercise and the cognitive interviews will be conducted individually per participant, not in groups.

Once we transcribed the cognitive interviews we then plan to have two independent assessors rate each participant's recollection against the ground truth in form of the full meeting transcript (as captured from the video recording). Possible rating dimensions are: the number of recalled issues; the order in which they have been discussed; details recalled for each discussed issue. The output will be a numerical *memory score* reflecting the accuracy and quality of recall over both conditions A and B. We will compare ratings by the two independent assessors using common inter-rater reliability measures.

Acknowledgments

The authors acknowledge the financial support of the Future and Emerging Technologies (FET) programme within the 7th Framework Programme for Research of the European Commission, under FET Grant Number: 612933 (RECALL).

REFERENCES

1. David M. Blei and John D. Lafferty. 2006. Dynamic Topic Models. In *Proceedings of the 23rd International Conference on Machine Learning (ICML '06)*. ACM, New York, NY, USA, 113–120. DOI: http://dx.doi.org/10.1145/1143844.1143859
2. David M. Blei, Andrew Y. Ng, and Michael I. Jordan. 2003. Latent Dirichlet Allocation. *J. Mach. Learn. Res.* 3 (March 2003), 993–1022. http://dl.acm.org/citation.cfm?id=944919.944937
3. N. Davies, A. Friday, S. Clinch, C. Sas, M. Langheinrich, G. Ward, and A. Schmidt. 2015. Security and Privacy Implications of Pervasive Memory Augmentation. *IEEE Pervasive Computing* 14, 1 (Jan. 2015), 44–53. DOI: http://dx.doi.org/10.1109/MPRV.2015.13
4. Seth Hunter, Pattie Maes, Stacey Scott, and Henry Kaufman. 2011. MemTable: An Integrated System for Capture and Recall of Shared Histories in Group Workspaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. ACM, New York, NY, USA, 3305–3314. DOI: http://dx.doi.org/10.1145/1978942.1979432
5. Alejandro Jaimes, Kengo Omura, Takeshi Nagamine, and Kazutaka Hirata. 2004. Memory Cues for Meeting Video Retrieval. In *Proceedings of the the 1st ACM Workshop on Continuous Archival and Retrieval of Personal Experiences (CARPE'04)*. ACM, New York, NY, USA, 74–85. DOI: http://dx.doi.org/10.1145/1026653.1026665
6. Chong Wang, David Blei, and David Heckerman. 2012. Continuous Time Dynamic Topic Models. *arXiv:1206.3298 [cs, stat]* (June 2012). http://arxiv.org/abs/1206.3298 arXiv: 1206.3298.
7. Xing Wei and W. Bruce Croft. 2006. LDA-based Document Models for Ad-hoc Retrieval. In *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '06)*. ACM, New York, NY, USA, 178–185. DOI: http://dx.doi.org/10.1145/1148170.1148204
8. Katrin Wolf, Yomna Abdelrahman, David Schmid, Tilman Dingler, and Albrecht Schmidt. 2015. Effects of Camera Position and Media Type on Lifelogging Images. In *Proceedings of the 14th International Conference on Mobile and Ubiquitous Multimedia (MUM '15)*. ACM, New York, NY, USA, 234–244. DOI: http://dx.doi.org/10.1145/2836041.2836065