Smile to save it: facial expressions for lifelogging

Citation
Venesvirta, H., Špakov, O., Gizatdinova, Y., Tuisku, O., Rantanen, V., Verho, J., ... Surakka, V. (2017). Smile to save it: facial expressions for lifelogging. In *Proceedings of the 16th International Conference on Mobile and Ubiquitous Multimedia* (pp. 441-448). ACM. https://doi.org/10.1145/3152832.3156628

Year
2017

Version
Peer reviewed version (post-print)

Link to publication
TUTCRIS Portal (http://www.tut.fi/tutcris)

Published in
Proceedings of the 16th International Conference on Mobile and Ubiquitous Multimedia

DOI
10.1145/3152832.3156628

Smile to Save It – Facial Expressions for Lifelogging

Hanna Venesvirta, Oleg Špakov, Yulia Gizatdinova
University of Tampere
Tampere, FI-33014, Finland
hanna.venesvirta@uta.fi
oleg.spakov@uta.fi
julia.kuosmanen@sis.uta.fi

Outi Tuisku
Lappeenranta University of Technology
Lahti, FI-15140, Finland
outi.tuisku@lut.fi

Ville Rantanen, Jarmo Verho
Tampere University of Technology
Tampere, FI-33101, Finland
ville.rantanen@tut.fi
jarmo.verho@tut.fi

Akos Vetek
Nokia Labs
FI-02610 Espoo, Finland
akos.vetek@nokia.com

Jukka Lekkala
Tampere University of Technology
Tampere, FI-33101, Finland
jukka.lekkala@tut.fi

Veikko Surakka
University of Tampere
Tampere, FI-33014, Finland
veikko.surakka@uta.fi

CONF '22, Jan 1 - Dec 31 2022, Authorberg.

unique doi string will go here

Abstract

Current aim was to introduce and initially evaluate the performance of a system called Extended Cognition, which was developed for enhancing the filtering of emotionally meaningful information from visual lifelogging data. The system is proposed to add facial expression markers to the stream of visual lifelogging data in order to later find special, possibly emotionally meaningful moments from the data. In an initial user study we collected subjective evaluations (*N*=10) about the ergonomics of the data measurement setup and the use of voluntary smiles and frowns for adding markers. Experience evaluations about the system were also collected. Results showed that data measurement setup was ergonomic to use and the system was evaluated positively. Smiling was rated as pleasant, natural, and functional expression for data marking purposes. Participants considered that, for example, the system would support memorization, and that it could be used to share important memories with others.

Author Keywords

Lifelogging; expression detection; affective computing.

ACM Classification Keywords

H.5.2. User Interfaces: Input devices and strategies. J.4. Social and Behavioral Sciences: Psychology.

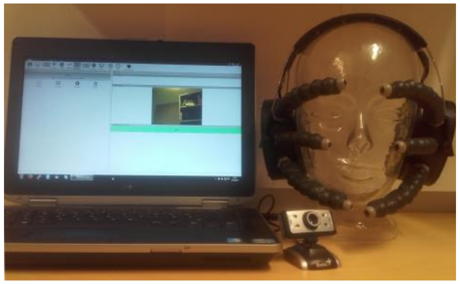

Figure 1. The EC data measurement setup: Wireless prototype detects facial expressions noninvasively. Any usb-connected common web-camera can be used for visual data capture. EC application is run currently on PC and it can be used both for data measurement (as shown in the figure) and for browsing the data.

Introduction

Lifelogging means storing digital data such as photographs, videos, physiological and behavioral data, or contextual information captured usually with wearables as memories of one's life [6,7,11,14,15,16]. In visual lifelogging, video and picture data usually shows first-person view of one's surroundings. As the continuous streaming of data heavily increases the amount of data, finding the most meaningful moments becomes a challenge. The visual data could be managed by utilizing the simultaneously recorded physiological or behavioral signals or contextual information as markers for the data [6,7,14,15,16].

Possibility to save and later return to emotionally meaningful moments and get easy access to emotional information is highly desired to function in the context of lifelogging [2]. Because facial expressions reflect socio-emotional information well [3], those could be utilized for adding emotional markers to the data stream [15,16]. This results that the emotionally meaningful moments can be easily restored from the data stream.

We propose Extended Cognition (EC) system, which takes an advantage of adding markers by facial expressions to the stream of visual lifelogging data. These markers are later used to find special moments from the data stream. Our main interest was to gain users' evaluations and opinions about the system and the proposed idea in the early development stage with a proof-of-concept implementation. Thus, current aim was to introduce the system and evaluate its performance in an initial user study by subjective assessments.

Extended Cognition system

With the proposed system, facial expression markers are added to visual lifelogging data in order to later find emotionally meaningful life moments based on the markers. The use of facial expressions for data marking require technology for facial behavior measurement, such as electromyography [4,5,21,22], capacitive measurement [19,20,21,23,24], or computer vision [9,14,15,16,25]. Capacitive measurement is noninvasive and easily multichannel as well as mobile measurement technique, which does not require attaching measurement sensors to the skin, nor it is vulnerable to changes in environmental lighting. Expression detection is based on the detection of the change of distance between capacitive sensor and skin. When an expression occurs, the distance between the skin and the measuring sensor changes, and the measured capacitance increases or decreases, and thus, an expression is detected [19,20].

The proposed system utilizes a proof-of-concept prototype (see Figure 1) with capacitive measurement for expression detection. The measuring sensors are located in six extensions attached to the earmuffs of a headphone-like device [19]. In the current study, controlled, voluntary facial expressions were used to produce high intensity expressions in order to guarantee good signal quality, which in turn provides robust expression detection. For the current study we selected smiling and frowning for data marking because they are commonly used in human communication, also voluntarily, and are also known to function well in human-computer interaction for input tasks [9,22,23,24]. Further, smiling and frowning related muscle activities are known to be related to positive and negative emotional experiences, respectively [12],

Figure 2. The data browsing view and the data browsing tools of the EC application. 1) The user can filter the pictures based on the expression marker. In the figure, the smile marker has been selected. 2) On the bottom of the application window, the filtered moments are shown as picture piles. 3) Piles are used to represent distinct and different lengths of moments. 4) The pictures from selected pile can be played as a slide show.

and thus, these can be used for marking moments with different emotional meanings. Visual data is recorded as still pictures at a preset rate with a common web-camera.

EC application is currently run on PC, and it can be used both for measuring and browsing the data. When the user browses the collected lifelogging data, they can go through either the whole picture database or use expression markers to browse only those pictures, which are associated with certain expression (see Figure 2). When an expression marker is used, pictures are automatically piled according to the timing and duration of the original marker. The content of a moment can be viewed as a slide show.

User study

Participants

Ten voluntary participants (6 females, mean age 28.7 years, range 22-36) took part in the experiment.

Apparatus

The EC system with the described wireless, wearable prototype device [19] and a common web-camera with 640 × 480 pixel resolution were used for data marking tasks.

A laptop PC ran EC application, which recorded one picture per second. Simultaneously, the application detected and recorded smiling and frowning events from the capacitive data. The detection of each expression was based on comparison of a signal from a sensor most sensitive to this expression against its baseline (BL) using a user dependent threshold (TH). TH was set manually based on a calibration, during which the participant produced controlled, voluntary expressions that caused visible changes in the capacitive signal. Based on the observed change in the signal, TH was set to be about one-third of the maximum signal peak. For calculating signal (dynamic) BL, a data buffer of 10 data points with a sampling rate of 40 Hz was used. BL was calculated as follows:

```
when new data point arrives
  add new data point to buffer
  if buffer is full
    if BL is not set
      BL = average (buffer)
    else
      avg = average (buffer)
      difference = abs (avg – BL)
      if difference < TH
        BL = BL + difference, max 0.05TH
    clear(buffer)
```

If the signal change exceeded TH, an event was detected and recorded to a data base. However, events shorter than 1.5 seconds were filtered out, as longer voluntary expressions were instructed. Afterwards, the collected pictures were filtered as piles of marked pictures based on the expressions captured at the time of recording.

Procedure

The study was divided in two phases. At the beginning of the first phase, after the participant signed a written consent form and provided background information, the EC measurement setup was worn. The measurement sensors were adjusted to detect frowning and smiling expressions and the expression detector was calibrated. The web-camera was positioned so that when the participant was standing in front of an object of interest

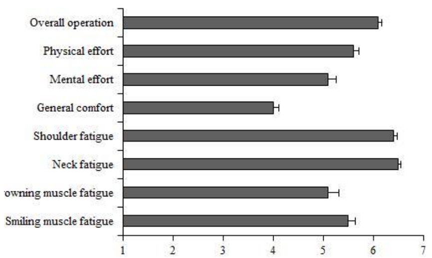

Figure 3. The ISO 9241-9 ratings.

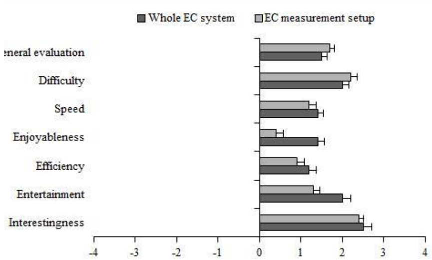

Figure 4. The experience evaluations of the use of the EC measurement setup and the whole EC lifelogging system.

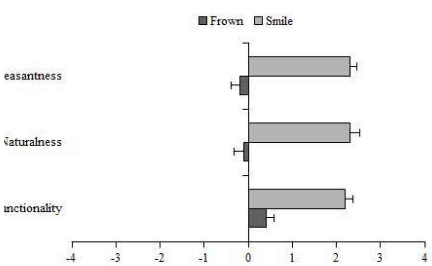

Figure 5. The ratings of the facial expressions.

viewing it, the object would be about in the middle of the resulting picture.

During a walk around in a building, the participants' task was to look for any objects of their own interest and mark the objects by either frowning or smiling based on whichever expression suited better for marking the object in case. Length for a marker was instructed to be about three seconds to guarantee high intensity expressions for robust detection. Participants were instructed to perform about 30 markers in total, about 15 per expression. The amount was marked down by the experimenter so the participant did not need to keep a record themselves.

After the marking, participant rated the use of the EC measurement setup with three rating forms. First, the ergonomics of the use of the measurement setup was rated using slightly modified ISO 9241-9 rating scales. The scales varied from 1 (e.g., very high physical effort) to 7 (e.g., very low physical effort) (see Figure 2) [10]. Second, evaluations about the use of facial expressions for marking were collected with three nine point bi-polar rating scales, which varied from -4 (e.g., unpleasant) to +4 (e.g., pleasant), and 0 represented neutral evaluation (see Figure 4) [9]. Finally, experience evaluations were given with seven bi-polar rating scales, which again varied from -4 (e.g., difficult) to +4 (e.g., easy) with 0 as neutral (see Figure 3) [9,22,23,24].

The second phase took place during the following week. The participants browsed the picture data they had collected during the first phase and tested the expression markers for filtering the data. These were done to demonstrate that already short data recording duration results in high amount of pictorial data, and further, to demonstrate the benefit of the expressions markers for data browsing. After the participant had tried the application, the EC lifelogging system as a whole was rated with experience evaluations. Finally, participant filled an end-questionnaire covering following themes: does the EC system have potential for the proposed lifelogging use; would the participants use the system; and how the participants considered the system could be developed further. With both phases, the study took about 45 minutes in total.

Results

The ratings for the ISO 9241-9 scales were positive throughout the scales (see Figure 3 for the mean ratings and the standard errors of the means (S.E.M.s)). Also all the ratings for the experience evaluation scales for the both phases were on the positive side of the scales (see Figure 4).

Figure 5 shows the mean ratings and S.E.M.s for the use of the facial expressions. Pairwise comparisons with Bonferroni-corrected Wilcoxon Signed-Rank tests showed that smiling was rated as significantly more pleasant ($Z = 2.34$, $p < 0.05$), more natural ($Z = 1.97$, $p < 0.05$), and more functional ($Z = 2.11$, $p < 0.05$) to use for data marking than frowning.

Nine of the participants thought the EC system has potential for the proposed lifelogging support. Seven participants reported that they would use EC system either as it currently is or after minor modifications. Eight participants said that the expression detection device should be lighter and less obtrusive. It was said that *"the system would help to notice which things were already forgotten"*, that is, the system could function as a memory support. It was also suggested that the system could function for marking already existing photographs or for measuring reactions during computer use. Most interestingly, it was said that the system could be used for sharing memories and their emotional meaning: *"[the EC system] gives a whole new way of sharing memories and makes it easier for others to really understand how it felt to be in a certain situation"*.

Discussion and future work

The aim was to introduce and evaluate EC lifelogging system, which can be used for adding facial expression markers to visual lifelogging data and later to find special, emotionally meaningful moments from the data. The results showed that the system has potential for the proposed purpose. First, the ISO 9241-9 ratings were all either on or above the middle value, indicating good ergonomics, both physical and mental, for data marking. Further, the experienced fatigue while using the system was quite low, which indicates that the system was considered quite effortless to use. This also applied to the use of voluntary expressions, even though these were instructed to be about three seconds in length. The ratings about the expression use also support this, as these were on the positive side for smiling and around the neutral value for frowning. The experience evaluations were all above the neutral value in both study phases, thus, the experience was quite positive. Participants clearly appreciated the idea and found it useful and entertaining. All the given positive evaluations indicate that the system functions well for adding facial markers and for using the markers in future filtering and retrieving of important and relevant data.

The experience evaluations in the Data browsing phase were generally more positive than the ratings given in the Data recording phase. It may be that after the participants had tested the EC application for data presentation and had become familiar with the whole EC system, they were able to see the full potential of the system. This was supported also through the end-questionnaire comments.

Smiling was rated as more natural, functional, and enjoyable to use for adding markers than frowning. These findings are in line with previous work: Ilves et al. [9] found that smiling was considered as more pleasant and somewhat more functional for game play purposes than frowning, and Tuisku et al. [23] found that smiling as a selection technique was evaluated with more positive ratings than other voluntary expressions, such as frowning. Here, smiling was also evaluated as more natural expression for adding expression markers than frowning. It is possible that making voluntary frowns is more complex than voluntary smiles, especially without any visual feedback about the expression. This may be due the fact that the lower face is more strongly represented on the primary motor cortex than the upper part of the face and thus, the voluntary control of lower face may be easier [13,18]. Further, it was spontaneously commented during post-experimental discussion that in a public place smiling may be more acceptable expression to use than frowning.

The participants considered that the EC system could function as memory support, it could support memorization, and it could save and restore emotion related moments. It is known that lifeloggers expect these all from their lifelogging systems [2]. Moreover, it was said that the EC system could function very well for sharing memories and their emotional meaning to other people. This would be an interesting future outcome, as it would enable that the EC system would also be evolved for emotion sharing and to function as a tool for technology mediated interaction.

The described initial study was conducted to demonstrate the performance of the EC system and to gain feedback about the idea. Despite the short duration and low sample size, the current results provide good basis for further development. In future, long term use of the system will be investigated. This is because it would be important to determinate that the current findings were not only because of the possible novelty effect or short duration of the use. For longer usage duration, less obtrusive design for facial expression detection device will be created, as the proof-of-concept prototype in its current form is relatively big in size. The data measurement with capacitive detection can be easily embedded to a smaller form such as near-eye-display, which would support less obtrusive measurement hardware design. However, we note that the current prototype device is about similar in size as commercially available head-mounted virtual reality displays, such as HTC Vive [8], and it is even smaller than virtual reality brain-computer interface (BCI) system added to HTC Vive [17]. Additionally, the behavior tracking methods the current prototype provides are far simpler than methods any BCI utilizes.

Current solution is limited to the detection of two voluntarily made expressions. Further development is needed to enable more meaningful functions of the system. First, in order to enable wider variety of expressions for marking moments with different emotional and cognitive meanings, the system requires more advanced expression detector, as introduced in Rantanen et al. [19]. In respect to detection of spontaneous expressions we are confident that capacitive detection would function well at least when the intensity of spontaneous expressions is comparable to current voluntary ones. This will be studied in future.

We will also investigate means for sharing the emotionally meaningful moments and memories, as proposed by the participants. Importantly, dynamic representations covering such things as expression duration and intensity will be used as, for example, those are more easily recognized than static representations [1].

In sum, we found that facial expressions can be used for enhancing the management of visual lifelogging data. The EC system satisfies several lifeloggers' needs, such as the possibility to save emotionally meaningful information and to reminisce and return to those moments [2]. In future, the system could also be used as a tool for emotion sharing, which in turn would support relatedness among people.

Acknowledgments

This work has been funded by the School of Information Sciences, University of Tampere, Finland; the Finnish Funding Agency for Technology and Innovation (Tekes), and Nokia Technologies. Authors want to acknowledge LUT research platform on Smart Services for Digitalisation.

References

1. Zara Ambarad, Jonathan W. Schooler, and Jeffrey F. Cohn. 2005. Deciphering the Enigmatic Face. The Importance of Facial Dynamics in Interpreting Subtle Facial Expressions. *Psychological Science* 16, 5: 403-410.
2. Yi Chen and Gareth J. F. Jones. 2012. What do people want from their lifelogs? In *Proceedings of the 6th Irish human computer interaction conference* (iHCI2012).
3. Paul Ekman. 1984. Expression and the nature of emotion. In *Approaches to emotion*, Klaus R. Scherer and Paul Ekman (Eds.). Lawrence Erlbaum Associates, Hillsdale, NJ, USA, 319-343.
4. Paul Ekman and Harriet Oster. 1979. Facial expressions of emotion. *Annual Review of Psychology* 30, 1: 527-554.
5. Alan J. Fridlund and John T. Cacioppo. 1986. Guidelines for human electromyographic research. *Psychophysiology* 23, 5: 567-589.
6. Jennifer Healey and Rosalind W. Picard. 1998. StartleCam: A cybernetic wearable camera. In *Proceedings of the 2Nd IEEE International Symposium on Wearable Computers* (ISWC), 42-49. https://doi.org/10.1109/ISWC.1998.729528
7. Steve Hodges, Emma Berry, and Ken Wood. 2011. SenseCam: A wearable camera that stimulates and rehabilitates autobiographical memory. *Memory* 19, 7: 685-696.
8. HTC Vive. Retrieved October 13, 2017 from https://www.vive.com/eu/
9. Mirja Ilves, Yulia Gizatdinova, Veikko Surakka, and Esko Vankka. 2014. Head movement and facial expressions as game input. *Entertainment Computing* 5, 3: 1875-9521.
10. ISO 9241-9:2000. 2000. Ergonomic requirements for office work with visual display terminals (VDTs) - Part 9: Requirements for non-keyboard input devices. CEN.
11. Tim Jacquemard, Peter Novizky, Fiachra O'Brolcháin, Alan F. Smeaton, and Bert Gordijn. 2014. Challenges and opportunities of lifelogging technologies: A literature review and critical analysis. *Science & Engineering Ethics* 20, 2: 379-409.
12. Jeff T. Larsen, Catherine J. Norris, and John T. Cacioppo. 2003. Effects of positive and negative affect on electromyographic activity over zygomaticus major and corrugator supercilii. *Psychophysiology* 40, 5: 776-785.
13. Robert W. Levenson, Paul Ekman, and Wallace V. Friesen. 1990. Voluntary facial action generates motion-specific autonomic nervous system activity. *Psychophysiology* 27, 4: 363-383.
14. Daniel McDuff, Amy Karlson, Ashish Kapoor, Asta Roseway, and Mary Czerwinski. 2012. AffectAura: An intelligent system for emotional memory, In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '12), 849-858. https://doi.org/10.1145/2207676.2208525
15. Atsushi Morikuni, Hiroki Nomiya, and Teruhisa Hochin. 2012. Taking lifelog videos and managing impressive scenes. In *Proceedings of the 13th ACIS international conference on software engineering, artificial intelligence, networking and parallel/distributed computing* (SNPD), 791-796. https://doi.org/10.1109/SNPD.2012.57
16. Atsushi Morikuni, Hiroki Nomiya, and Teruhisa Hochin. 2015. Expression strength for the emotional scene detection from lifelog videos, In *Software Engineering Research, Management and Applications*, Roger Lee (Ed.). Springer International Publishing, Berlin, Heidelberg, Germany, 127-140.
17. Neurable. Retrieved October 13, 2017 from http://www.neurable.com/

18. Wilder Penfield and Theodore Rasmussen. 1950 *The cerebral cortex of man: A clinical study of localization of function*. Macmillan, New York, NY, USA.

19. Ville Rantanen, Pekka Kumpulainen, Hanna Venesvirta, Jarmo Verho, Oleg Špakov, Jani Lylykangas, Akos Vetek, Veikko Surakka, and Jukka Lekkala. 2013. Capacitive facial activity measurement. *ACTA IMEKO* 2, 2: 78-85.

20. Ville Rantanen, Pekka-Henrik Niemenlehto, Jarmo Verho, and Jukka Lekkala. 2010. Capacitive facial movement detection for human-computer interaction to click by frowning and lifting eyebrows. *Medical and Biological Engineering and Computing* 48, 1: 39-47.

21. Veikko Surakka and Jari K. Hietanen. 1998. Facial and emotional reactions to Duchenne and non-Duchenne smiles. *International Journal of Psychophysiology* 29, 1: 23-33.

22. Veikko Surakka, Marko Illi, and Poika Isokoski. 2004. Gazing and frowning as a new human-computer interaction technique. *ACM Transactions to Applied Perception* 1, 1: 40-56.

23. Outi Tuisku, Ville Rantanen, Oleg Špakov, Veikko Surakka, and Jukka Lekkala. 2016. Pointing and selecting with facial activity. *Interacting with Computers* 28, 1: 1-12.

24. Outi Tuisku, Veikko Surakka, Toni Vanhala, Ville Rantanen, and Jukka Lekkala. 2012. Wireless Face Interface: Using voluntary gaze direction and facial muscle activations for human-computer interaction. *Interacting with Computers* 21, 1: 1-9.

25. Zhihong Zeng, Maja Pantic, Glenn I. Roisman, and Thomas S. Huang. 2009. A survey of affect recognition methods: audio, visual, and spontaneous expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 31, 1: 39-58.