

Reproduction of Internal States by Video-Stimulated Recall of Enjoyable Activities

Kanako OBATA¹, Hirotada UEDA¹, Kei SHIMONISHI¹, Kazuaki KONDO¹,
and Yuichi NAKAMURA¹[0000–0001–5103–7194]

Academic Center for Computing and Media Studies, Kyoto University,
Sakyo, Kyoto, Japan {obata,yuichi}@media.kyoto-u.ac.jp

Abstract. Engaging in enjoyable activities with intimate persons have positive effects on both the physical and mental states of a person. For instance, co-eating or chatting with family members or friends tends to elicit positive emotions and feelings. This research aims to analyze the effects of recalling enjoyable experiences, and to gain insights into developing methods for effectively reproducing positive internal state. For this purpose, we conducted preliminary experiments in which original activities were recorded by videos and a heartbeat monitor, and these videos were presented to some participants after a certain number of days with facial expressions and heartbeats measured in the same ways as the original activities. The results show that facial expressions and heart rate features effectively characterize the internal states in the original experiences and when reviewing the recordings. Moreover, the results suggest that there are conditions for evoking internal states similar to the original experiences.

Keywords: reproduction of internal states · video recall · heart rate · smile

1 Introduction

When we look at photos or videos of our activities with friends or intimate people, we typically feel happy and at peace. The effect is presumably provided by the mechanism by which recall of our experiences significantly induces our emotions and feelings. This research focuses on measuring the effects of the mechanism; to what extent can videos or photos can evoke what types of effects.

This research is motivated by severe social demands for monitoring and preserving mental health from depression, isolation, frail, etc. Reminiscence therapy, one of the dementia treatments, utilizes the effects of early recollection of memories[1]. It aims to improve psychological well-being by retrieving self-efficacy, self-esteem, and the importance of existence by recollecting the early days, typically days at a young age.

One type of the activities we are focusing on is “eating with intimate people,” i.e., co-cooking or co-eating. These activities undoubtedly benefit our mental health, not only during the time of activities but also when we reflect on these

experiences. Thus, we can anticipate that providing photos or videos records of those activities will have a positive impact.

To analyze those effects, first, we need to measure the internal states both during actual experiences and when recalling them, and then analyze the recalled experiments by comparing the measuring results. For this purpose, we use facial expressions and heartbeats, and confirm the conditions for good effects. Second, we need to determine which portions of the records are more effective to draw good response, i.e., good internal states [3] as stated in episodic memory recall[2]. With these analyses, we can effectively design the use of activity records.

As preliminary experiments, we recorded chatting, co-cooking and co-eating activities and reviewed and analyzed the effects. Thorough experiments, we obtained good clues that support the above framework of inducing good internal states, and obtained some ideas for conditions under which this mechanism works effectively.

2 Activity and Internal States

2.1 Enjoyable Activities

We focus on the activities of chatting, cooking, and eating with intimate persons. We have an enjoyable time chatting and eating special meals together at Christmas time, the new year's day, and other festival seasons. Memories of those experiences are typically recalled triggered by similar dishes or photos, and we retrieve a happy feeling at those activities. This suggests that watching or hearing enough records of those activities would induce good internal states. Hereafter, this process of watching to recall experiences will be referred as the "reviewing process."

To use review processes, we need to consider negative conditions that those records do not provide good effects, e.g., photos or videos are less stimulus and boring, or sometimes even embarrassing. We also need to consider the situation in which a person does not have enough time to watch the records. Considering those backgrounds, we investigated which part of activities (records) cause good effects to the internal states.

2.2 Relationship of internal states between original expedience and reviewing process

We focus on positive internal states, e.g., happy, pleased, delighted, etc. that are most desirable emotions for maintaining mental well-being. Hereafter, we represent them by "positive state". This positive state can be specified by positive valence, and the rightmost area in Russel's circumplex model[4] may represent this state. This area typically causes facial expressions such as smiling, which is occasionally superseded by laughing.

The area does not correspond to a high arousal value in the circumplex, which means that the sympathetic nervous system is not highly active, and that the

parasympathetic nervous system is more active than it. This state is considered free from severe stress and strain, and is rather related to a relax mood or feeling. The state is partially measured by heart rate features.

From the above consideration, we employed smile ranking through activities and a video-reviewing process. Smile is easily recognized by humans and possibly by computer vision. We first attempted human rating; however, we are considering automatic smile rating in the future[6][7]. The other measurement is on the heartbeat. Some feature indices of heartbeat e.g., LF/HF, SDNN, and RMSSD, are closely related to the activation of the sympathetic nervous and the parasympathetic nervous systems. The heart rate features are supposed to provide supplemental information to smile ranking by the estimation of a stressed or relaxed state.

The following important points should be clarified:

- How can we reliably measure positive states using facial expressions and heart rate features?
- What are the internal state relationships between the original experience and the reviewing process?
- What are the individual differences among the participants?
- How does the reviewing condition (including the internal state) affect the reviewing results?
- How long does the effect last?

Each of the above points includes a variety of factors, and we first deal with the first and second points in this article.

3 Measurements

We record the states of the participants using videos and a heart rate monitor during original activities and reviewing processes.

3.1 Recording method

For video recording, we mainly use omnidirectional cameras (GoProMAX) to capture the faces of the participants and important objects such as dish and drink, as shown in Figure 1 (a). If necessary, two or more cameras were placed to capture the faces of the participants and environments. Speeches and other important sounds are recorded and included in the videos. Figure 1 (b) shows the recording method for video reviewing. An omnidirectional camera is placed to capture the face of the participant and video content.

Heartbeats are recorded using wearable heart rate monitors (myBeat WHS-1) during both the original activities and reviewing processes. These sensors (monitors) also record acceleration in three dimensions, which is used for data synchronization with the video data¹.

¹ The wearing participant jumps while being captured by the video



Fig. 1. (a) Original activity and video capturing. Omnidirectional camera(s) is placed to capture the faces of the participants and the environment. (b) Video-reviewing process. Omnidirectional camera(s) is placed to capture the face of the participant and the video monitor.

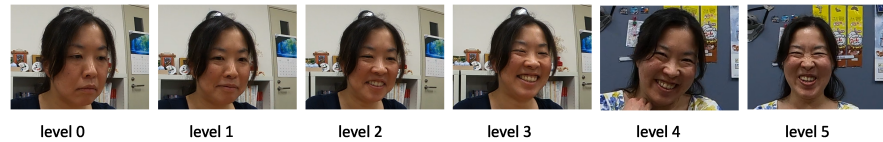


Fig. 2. Example of smile rating. Facial expressions are classified -1 (negative), 0 (neutral) to 5 (smile with the highest intensity)

3.2 Smile index

Facial expressions in reviewing video records potentially represent the internal states[5]. For this purpose, a human rater scores facial expressions every second ranging from ?1 to 5, with “0” representing neutral, “5” representing the maximum smile, and ?1 representing negative. Figure 2 shows sample images of facial expressions.

The rating was done by a single person in the following experiments; however, multiple raters will be preferable for future experiments[6][7]. In the current experiments, we do not detect and remove fake smiles, and this is left for future work. Rating results are processed by moving averages to reduce the effects of fluctuations and quantization errors. We use the resultant sequence of the averaged ranks as “smile-index,” To analyze the relationships of two facial expression sequences, e.g., person A and B, we calculated the correlation between two sequences of facial expression changes for 10 s.

3.3 Heart rate variability

RRI (interval between R waves) has been intensively analyzed. The high-frequency component of RRI (HF) is mainly related to the parasympathetic nervous systems; the low-frequency component of RRI (LF) is related to both the sympathetic and parasympathetic nervous systems. However, because RRI is significantly affected by body movements and other artifacts, LF/HF is commonly used

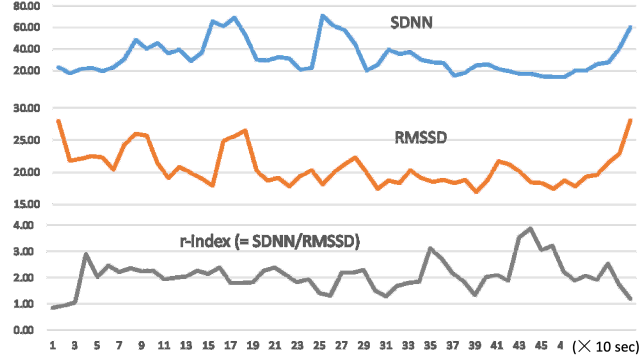


Fig. 3. Sample of heart rate variability (SDNN, RMSSD, r-index (=SDNN/RMSSD))

as an index of the activation of the parasympathetic nervous system. Thus, a smaller LF/HF value suggests a relaxed mental state, and a larger value suggests a stressed or strained internal state.

LF and HF are indices in the frequency domain, which cannot be reliably calculated in a short period. Alternatively, SDNN and RMSSD are frequently used as indices that are tightly related to the activation of the sympathetic and parasympathetic nervous systems [8][9]. They are calculated as follows.

$$SDNN = \sqrt{\frac{\sum_{i=1}^N (I_i - \tilde{I})^2}{N}} \quad (1)$$

$$RMSSD = \sqrt{\frac{\sum_{i=2}^N (I_i - I_{i-1})^2}{N - 1}} \quad (2)$$

where I_i represents the i -th RRI of heartbeat, and $\tilde{I}$ represents the average of RRIs.

RMSSD is considered the index of the superiority of the parasympathetic nervous system to the sympathetic nervous system, which may correspond to the HF element. SDNN is related to the activation of both nervous systems. These indices are also significantly affected by body movements, postures, and speaking acts. The ratio of SDNN and RMSSD (SDNN/RMSSD) is devised to cancel those artifacts by dividing SDNN by RMSSD, both of which values are increased by body movements[10][11]. Hereafter, we represent the ratio as “r-index” ($r = \text{SDNN}/\text{RMSSD}$).

It is common to calculate SDNN and RMSSD for 3- to 5-min epochs, and the person being measured is typically requested to stay still (at rest) during the measurement. Our purpose is to measure the internal states of a person while he/she is engaged in activities, such as chatting, cooking, and eating. The internal state in such situations is expected to change more quickly than the above resting case. To deal with this condition, we calculate these indices every

Table 1. The idea of categorizing internal states by facial expression and heart rate features. P denotes “positive,” R denotes “relaxed,” S denotes “stressed,” m denotes “possible effects of movement,” and $*$ denotes “uncertain,” that is, “cannot be estimated.”

smile-index	small r-index		large r-index	
	large SDNN	small SDNN	large SDNN	small SDNN
large smile-index	$P \oplus Rm$	$P \oplus R$	$P \oplus m$	$P \oplus S$
small smile-index	$* \oplus Rm$	$* \oplus R$	$* \oplus m$	$* \oplus S$

30-s epoch. Figure 3 shows a sample in our experiments. SDNN, RMSSD, and r-index (=SDNN/RMSSD) changed based on both body motions and internal states, and this does not show unacceptable fluctuations. We intensively analyze the proper use of the measurement of these indices in a short epoch such as a 30-s epoch.

3.4 Analysis using two modalities

We analyze videos and heartbeats that are simultaneously recorded in the original activities and their reviewing processes. The meaning of heart rate indices is partially deduced from video recordings by examining motion and speech conditions.

Considering the characteristics of each feature index, we have following classification of internal states as shown in Table 1

- Large smile-index represents positive internal states. However, it does not always express relaxed states. We need to assess heart rate features to clarify it.
- Low r-index suggests relaxed internal states.
- If the r-index is large, we need to check the SDNN value. If the SDNN value is large, we assume that this portion expresses body movements and does not provide clues for a stressed or relaxed state. If the SDNN value is small, it suggests a stressed internal state.

We can use SDNN as a representative of SDNN and RMSSD because both simultaneously become large with body movements. If the SDNN (or RMSSD) value is small, it implies small and negligible body movements. Consequently, a state with a large r-index and small SDNN is considered stressed or strained. In contrast, a state with a large r-index and SDNN does not provide clues for either a stressed or relaxed state, because a large SDNN does not necessarily mean body movements. It is assumed that a small r-index value represents a low LF/HF index and indicate relaxed rather than stressed states.

Hereafter, the feature values f of the i -th participant in the original experience and a reviewing process are denoted as $f_i^o(t)$ and $f_i^r(t)$, respectively. t represents the timeline of the original experience and its record. f can be smile-index (Si), r-index (Ri), SDNN, RMSSD, etc. $\tilde{f}_{i \neq k}^o(t)$ represents average of f^o for all participants, except the k -th participant.



Fig. 4. Experiment 1: chatting with an intimate friend

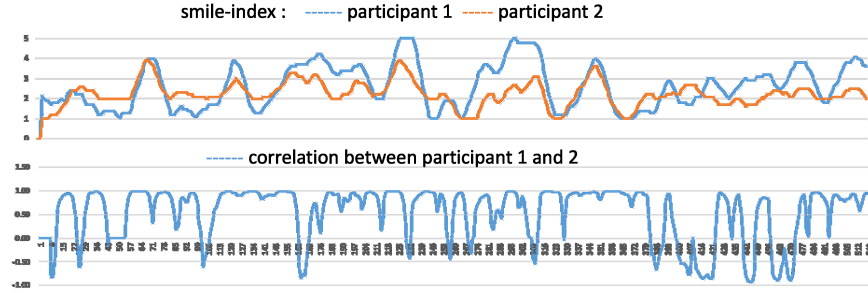


Fig. 5. Smile-index changes in experiment 1

4 Preliminary Experiments

4.1 Recorded activities

We recorded the following activities.

Experiment 1 Figure 4 shows sample images of the activity of chatting with an intimate friend. We did not notice any significant body movements, such as standing up or walking, from the participants while they sat in front of the table. One participant wore a heart rate monitor. The length of chatting was approximately 25 min. The reviewing process was done by one of the participants alone without communicating with others, as shown in Figure 1 (b). The reviewing process was performed 3 weeks after the original activity.

Experiment 2 Four participants in the same office cooked pancakes together, as shown in Figure 1 (a). Throughout the activity, they chatted, cooked, and ate pancakes, the effects of which were recorded. Each participant wore a heart rate monitor. Two of the participants reviewed the video recording. The reviewing process was performed approximately 5 weeks after the original activity. All reviewing processes were performed by one participant alone without communicating with others.

4.2 Smile and heart rate features

First, we show the measurement results and the analysis results of the original activities. The upper row of Figure 5 shows smile-index changes through Ex-

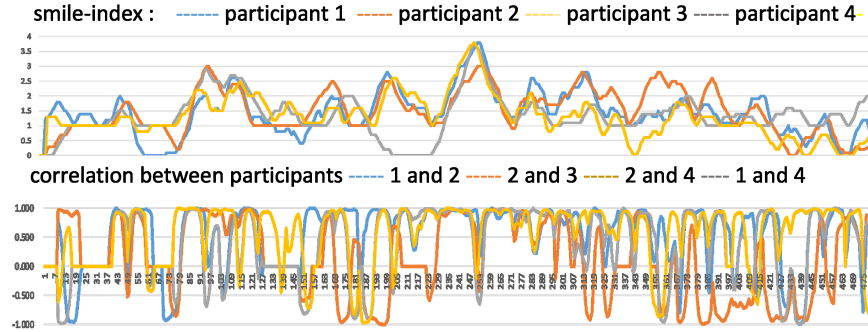


Fig. 6. Smile-index changes in experiment 2

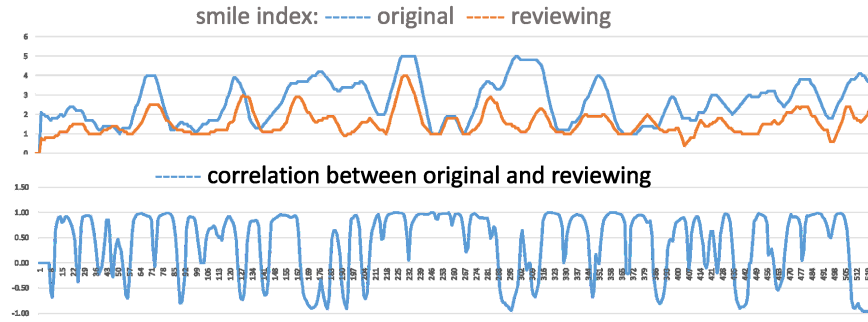


Fig. 7. Smile-index changes during the reviewing process in Experiment 1

periment 1. As shown in the figure, the smile-indices of both participants are relatively high on average, and peaks intermittently appeared. The correlation between the smile-indices of the two participants throughout the experiment is approximately 0.6, as described in the lower row of Figure 5. It is mostly positive, which implies that the chatting activity was fun for both participants and that they felt sympathy.

The upper row of Figure 6 shows smile-index changes through Experiment 2. Similar to Experiment 1, smile-index peaks appear intermittently; however, the intensity is relatively low compared with Experiment 1. The lower row of Figure 6 shows detailed correlation between participants. The correlation shows more complicated changes in which negative values frequently appeared, presumably due to sophisticated communications among four participants while attention was frequently paid to the cooking tasks such as success or failure of each process.

Figure 7 shows the smile-index during the reviewing process in Experiment 1. In the original activity, body motions significantly affected the indices; however, in a reviewing process, body movements are negligible, and the indices are expected to represent mostly internal states. The lower row of the figure depicts the

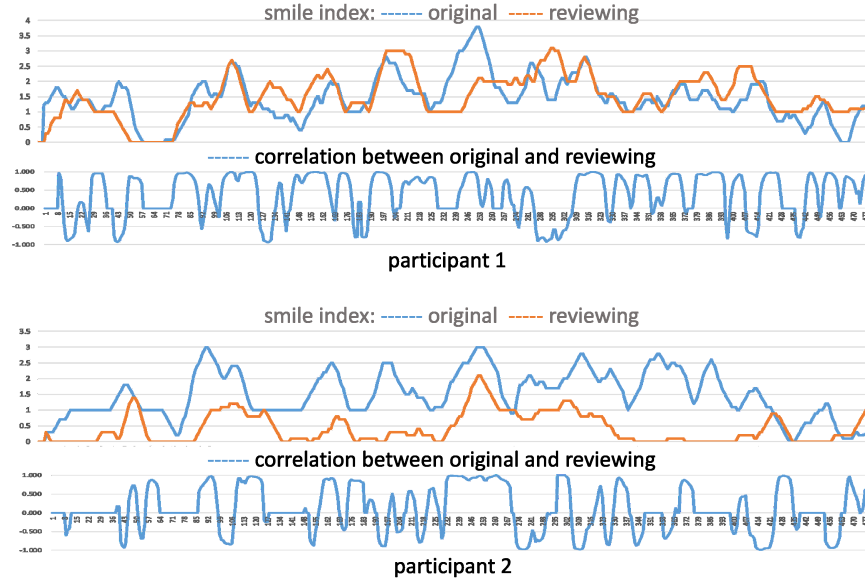


Fig. 8. Smile-index changes during the reviewing processes in Experiment 2

correlation to the original activity. Mostly, similar facial expressions appeared as the original activity even though some portions have a lower correlation, and the reason for this is discussed below.

Figure 8 shows the smile-index during the reviewing process in Experiment 2. The baseline of the intensity is stable and lower than that of the original activity, especially for participant 2; however, there are some significant peaks almost aligned with the peaks of the original activities.

Figure 9 and 10 show the heart rate features through Experiments 1 and 2. The figures show that SDNN and RMSSD have almost the same trend and peak positions, with a correlation of approximately 0.9. This supports the above discussion in Section 3.4 that we can use either of them to detect trends and peaks. Different from those indices, r-index has a lower baseline and intermittent spikes that suggest the activation of the sympathetic nervous system. In the experiments, many of those peaks correspond to situations with significant body movements or intensive chatting.

Each SDNN, RMSSD, and r-index correlation between participants is low, with the majority of them being close to zero. Examples of details are shown in Figure 9(b) and Figure 10(b). Positive and negative portions alternated, implying that raw heart rate features are not systematically synchronized.

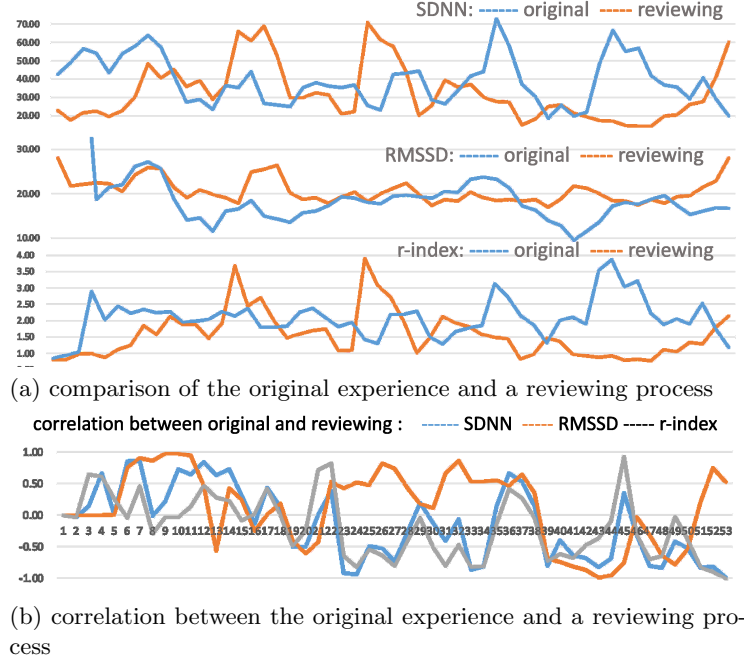


Fig. 9. Heart rate features in Experiment 1

4.3 Relationship between the original and reviewing

Before discussing the relationships, we first categorize the situations in the original activities. The following are four typical situations that cause peaks (local maxima) of smile-index or local extrema of r-index.

Suppose that we are focusing on one participant, denoted by X , and others are denoted by Y_i .

- (A) body movements or discussion of Y_i created a lively mood
- (B) body movements or discussion of X created a lively mood
- (C) body movements or discussion of X and one or more Y_i created a lively mood
- (D) body movements or discussion of X or Y_i had no significant effect on mood

As shown in the above examples, smile-index peaks in the original activities typically resulted in smile-index peaks in a reviewing process. Generally, smile-index has a significant correlation between the Original experience and a reviewing process, that is, denoted as $\text{correlation}(Si_i^o, Si_i^r)$. The lower parts of Figure 5 and Figure 6 show examples. Correlation between self smile-index in the original experience and that in a reviewing process ($\text{correlation}(Si_i^o, Si_i^r) = 0.59$ in Figure 5 case) is slightly higher than the correlation to other members

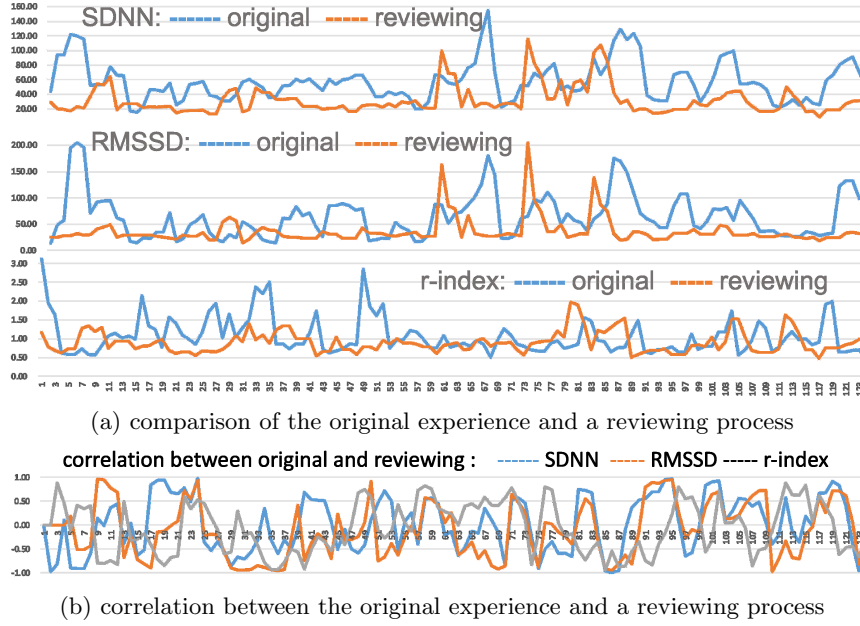


Fig. 10. Heart rate features in Experiment 2

($\text{correlation}(Si_{k \neq i}^o, Si_i^r) = 0.42$ in Figure 5 case). This suggests that smile reproduction strongly depends on the self internal states of the original experiment rather than the atmosphere in the original experiment.

More precisely, the above categories (A) and (C) are the most frequent Types, where larger peaks are reproduced in a reviewing process. Category (B) also occasionally reproduces smile-index peaks in a reviewing process; however, the frequency is less, and the peak values are typically lower than those of category (C). Smile-index peaks in a reviewing process always have corresponding peaks in the original experiences. We did not observe smile-index peaks without the corresponding peaks in the original experiences.

Looking at heart rate indices, the correlation between the original and reviewing process is approximately 0, which is similar to that between participants. However, some extrema indicating relaxed/stressed states in the original experience may have induced relaxed/stressed states during a reviewing process, as indicated by the red/green solid lines on the graph in Figure 11. These are typical examples of the above categories (A) and (C) for relaxed states, (D) for stressed or strained states, or unrelated to positive internal states. Counter samples that do not follow this idea are at the dotted orange lines in the same graph, and the situation corresponds to (B). This is probably because watching the participant's behaviors objectively is sometimes embarrassing, which may cause stress. One of the participants introspected that he/she felt embarrassed in watching his/her big smiles/laughs. In this regard, when one participant's behaviors amused the

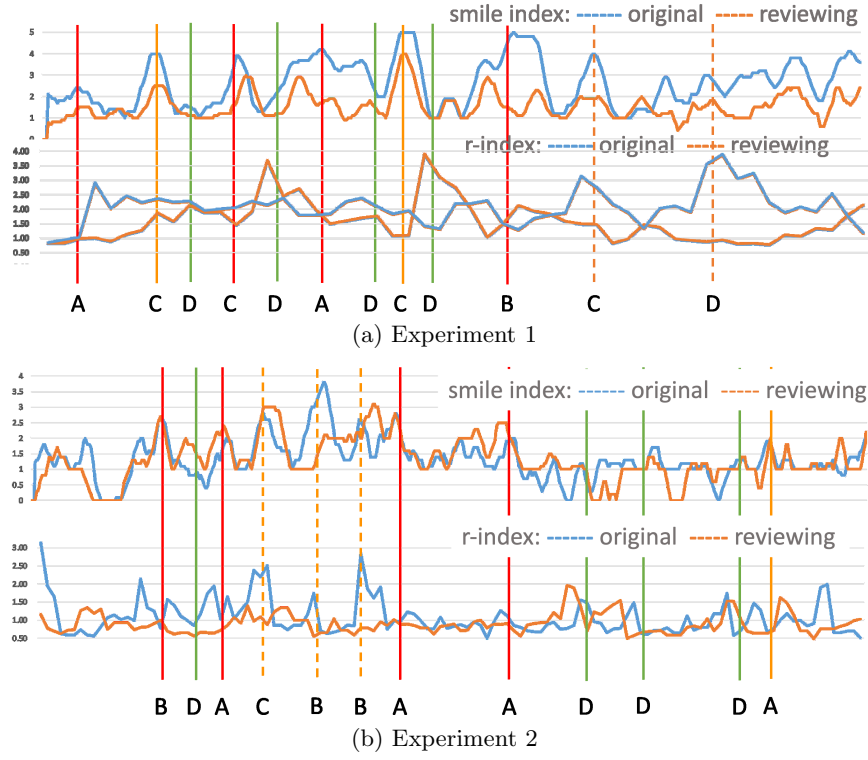


Fig. 11. Relationship on internal states between the original experience and a reviewing process in (a) Experiment 1 and (b) Experiment 2

others, the person and others may exhibit different trends. For example, the person has a higher r-index (stressed or strained), whereas the others have a lower r-index (relaxed).

Next, let us consider the mutual relationships between facial expressions and heart rate features. The red solid lines in Figure 11 correspond to $P \oplus R(m)$, and the green solid lines correspond to $* \oplus S$ given in Table 1. If those conditions are satisfied, the same internal states tend to be induced, as demonstrated in the example. The orange solid lines represent $P \oplus S$, which correspond to the above categories (B) or (C) in which the participant's actions elicited a positive mood and smiles. In such cases, smiles may also appear, but they tend to be weaker than in the typical cases of (C).

The orange dotted lines represent $P \oplus m$. Although smiles may appear in a reviewing process of such portions, smiles tend to be significantly weaker, and internal states show stressed states such as $P \oplus S$ or $* \oplus S$. Most of these types correspond to the cases that we already discussed in the above cases, that is, the participant was slightly embarrassed. Another possibility is that the participant

recalled the sophisticated situations, e.g., the participant pretended to smile (a wry smile) a little bit in the original experiment.

In addition, intense smiles frequently occur ($P \oplus Rm$) just after some stressed situations in which r-index is relatively high ($* \oplus S$). This can be also a good clue to finding portions that reproduce positive internal states.

5 Conclusion

In this article, we investigated the effects of recordings of enjoyable experiences to reproduce positive internal states. For the analyses, we recorded the chatting, co-cooking, and co-eating activities of participants and measured their smiles and heart rate features. Through our preliminary experiments, the smile-index correlation between the original activity and a reviewing process was kept relatively high, and parts with intense smiles in a reviewing process were frequently accompanied by relaxed internal states.

These experiments are still preliminary, and most of the analyses are only qualitative. We need a systematic investigation for quantitative analysis. With a more systematic understanding of the effects of reviewing processes, we can expect those effects for therapy, QOL monitoring, etc.

References

1. Y. Lin, Y. Dai, S. Hwang, “The Effect of Reminiscence on the Elderly Population: A systematic Review”, *Public Health Nursing*. 20 (4): 297–306, 2003
2. H. Le, S. Clinch, C. Sas, T. Dingler, N. Henze, N. Davies, “Impact of Video Summary Viewing on Episodic Memory Recall – Design Guidelines for Video Summarizations”, *CHI ’16*, 2016
3. E. Siedlecka, F. Denson, “Experimental Methods for Inducing Basic Emotions: A Qualitative Review”, *Emotion Review*, 11(1), 87–97, 2019
4. J. Russell. A circumplex model of affect. *Journal of Personality and Social Psychology*, Vol. 36, pp. 1161–1178, 1980
5. M. Haj, M. Daoudi, K. Gallouj, A. Moustafa, J. Nandrino, “When your face describes your memories: facial expressions during retrieval of autobiographical memories”, *Reviews in the Neurosciences*, 2018
6. J. Hideo, S. Jacopo, S. Nicu, J. Joemon, “Looking at the viewer: Analysing facial activity to detect personal highlights of multimedia contents”, *Multimedia Tools Appl.* Vol.51. pp.505-523, 2011
7. K. Kondo, T. Nakamura, Y. Nakamura, S. Sato, “Siamese-structure Deep Neural Network Recognizing Changes in Facial Expression According to the Degree of Smiling”, *The 25th Int’l Conference on Pattern Recognition (ICPR2020)*, 2021
8. H. Otzenberger, C. Gronfier, C. Simon et al., “Dynamic heart rate variability: A tool for exploring sympathovagal balance continuously during sleep in men”, *American Journal of Physiology*, vol. 275, no. 3, pp. H946–H950, 1998
9. R. Balocchi, F. Cantini, M. Varanini, G. Raimondi, J. Legramante, A. Macerata, “Revisiting the potential of time-domain indexes in short-term HRV analysis”, *Biomedizinische Technik*, vol. 51, no. 4, pp. 190–193, 2006

10. J. Sollers, T. Buchanan, S. Mowrer, L. Hill, and J. Thayer, "Comparison of the ratio of the standard deviation of the r-r interval and the root mean squared successive differences (SD/rMSSD) to the low frequency-to-high frequency (LF/HF) ratio in a patient population and normal healthy controls", *Biomedical Sciences Instrumentation*, vol. 43, pp. 158–163, 2007
11. Hui-Min Wang and Sheng-Chieh Huang, "SDNN/RMSSD as a Surrogate for LF/HF: A Revised Investigation", *Modelling and Simulation in Engineering*, Volume 2012, Article ID 931943, 2012