Transient Stress Stimulus Effects on Intentional Facial Expressions

- Estimation of Psychological States based on Expressive Tempos -

Kazuhito Sato
Department of Machine Intelligence and Systems Engineering,
Faculty of Systems Science and Technology, Akita Prefectural University
Yurihonjo, Japan
ksato@akita-pu.ac.jp

Takashi Suto
IAI Corp.
Mechanical Design Section 2
Shizuoka, Japan
takashi-suto@iai-robot.co.jp

Hirokazu Madokoro
Department of Machine Intelligence and Systems Engineering,
Faculty of Systems Science and Technology, Akita Prefectural University
Yurihonjo, Japan
madokoro@ akita-pu.ac.jp

Sakura Kadowaki
Smart Design Corp.

Akita, Japan
sakura@smart-d.jp

***Abstract*-This paper presents a framework of tempos and rhythms to clarify the relevance between psychological states and facial expressions, particularly addressing repetitive operations of intentional facial expressions after giving a stress stimulus. By acquiring image datasets of facial expressions under the states of pleasant–unpleasant stimulus for 20 subjects, we extracted expressive tempos for respective subjects. Consequently, averages of extraction rates show that the pleasant state was 81.1%. The unpleasant state was 77.8%. Regarding effects of pleasant–unpleasant stimulus on the expressive tempos, particularly addressing the variation of the number of frames constituting one tempo, the variation in unpleasant stimulus became greater than that in the pleasant stimulus. The results show that the analysis using expressive tempos and rhythms is valid as an indicator for estimating the psychological state.**

Keywords-Psychological measures, stress; Intentional facial expression; Machine learning approaches; Behavior modeling.

I. INTRODUCTION

Humans can feel rhythms from all of their personal surroundings that are moving, especially any emitting sound. Additionally, they feel rhythms from engaging in daily life, such as rhythms related to conversation and rhythms of human life [1][2]. Among these, biological rhythms are based on personal tempos. In other words, personal tempos are individual-specific, not derived from physiological functions. It has been reported that personal tempos also vary depending on environments and moods [1]. In daily life, for behaviors such as walking or talking, personal tempos represent individual-specific speeds, which are expressed naturally in a free-action situation without constraint. For a facial expression as a daily life behavior, we infer that individual-specific rhythms can exist also.

To clarify the relevance between psychological states and facial expressions, we propose a framework of rhythms and tempos that specifically examines actions to repeat intentional facial expressions after giving a stress stimulus. We define one rhythm as one tempo repeated several times. In addition, we regard one tempo as the period during which facial expressions transform from a neutral face (i.e., expressionless) to the next neutral face through the maximum number of facial expressions, in a time-series variation of Expression Levels (ELs), i.e., labels quantifying exposed levels from the neutral face [3]. We use Hidden Markov Models (HMMs) [4] of left-to-right type to extract expressive tempos. As a method to classify categories extracting the occurrence part of patterns from the time series data, HMMs are widely used in fields of signal processing and speech recognition. They can extract expressive tempos, which represent the occurrence pattern of exposed intensities. Stress reactions appear in relation to biological phases (e.g., changes in heart rate, changes in blood pressure), psychological phases (e.g., depression, irritability), and behavioral phases (e.g., increase of drinking and smoking, fidgety state) [5]. Here, the facial expression is classified as a behavioral phase among the stress reactions. For this reason, by analyzing the rhythm and tempos that appear after exposure to different stress conditions, we infer that the inference of psychological states, such as comfort

 ISBN: 978-1-61208-356-8

and discomfort, from the changes of individual-specific facial expressions will become possible in the future.

In this study, as the basis for objectively expressing the ambiguity and complexity of facial expressions attributable to the psychological stress states of human, we propose a framework of exposed rhythms and tempos on intentional facial expressions. This study might derive the following advantages in applications. One familiar case of those is to develop a training tool to create an attractive smile that hospitality mind is easily transmitted to the customer. Foreseeable future, we could be sure that this study is valid as new indices for detecting the distraction state of driver by time-series changes of eye-gaze and facial expressions.

This paper is presented as the following. We review related work to clarify the position of this study in Section II. In Section III, we define a new framework of exposed rhythms and tempos for analyzing relations of psychological stress and facial expressions. In Section IV, we describe the method to capture facial expression images, preprocessing, classification of facial expression patterns with self-organizing maps, integration of facial expression categories with fuzzy adaptive theory, extraction of expressive tempos using HMMs. We explain our originally developed facial expression datasets including stress measurements in Section V. In Section VI, we optimize the number of states of HMMs by extracting expressive tempos from facial expressions and .analyze the transient stress stimulus of pleasant–unpleasant effects on the expressive rhythm of facial expressions. Finally, we present conclusions and intentions for future work in Section VII.

II. RELATED WORKS

Open datasets [6][7][8]of facial expression images are released from some universities and research institutes to be used generally in many studies for performance comparisons of facial expression recognition or automatic analysis of facial expressions. These datasets contain a sufficient number of subjects as a horizontal dataset. However, images are taken only once for each person. As one of recent researches using these datasets, there is the study by Das and Yamada [9]. They used the Cohn–Kanade [6] and the Extended Cohn–Kanade (CK+) datasets [7] to obtain emotional mixture or percentage composition of emotion data, because cross-sectional datasets are valid rather than time-series datasets in evaluating stress. The CK+ datasets contain Action Units (AUs) coded facial image data with lead emotion label for each peak expression. Therefore, they considered the peak and few intermediate states of each facial expression taking care that the difference in intensities is not large enough to represent another emotion altogether. Das and Yamada conducted two moderate sized surveys to correlate individual emotions to stress and to find relationship between predicted emotional mixtures of facial expressions and stress levels [9]. After predicting emotional composition, they selected facial expression images for two surveys. However, the respondents were just only instructed to look at the static facial image and label the stress levels from 0 to 9 according to each individual perception. Consequently, Das and Yamada did not carry out analysis that focused on the expressive process of individual-specific facial expressions, in spite of lurking clue in there.

In a study particularly addressing the dynamic aspects of facial expressions, Hirayama et al. [10] found the kinetic period of face parts. They have proposed an expressive notation as a representative format that describes the timing structure on facial expressions. They were seeking linear systems (i.e., modes) to the bottom-up from feature vector sequences. The modes represent various motional states or stationary states of face parts. For example, in the case of the mouth, there are open, remain open, close, and keeping closed as elements of mode sets. The method explained by Hirayama et al. tracked feature points, i.e., a total of 58 points is assessed from the outline of the lower half face including each eyebrow, each eye, nose, and lips. Then using Active Appearance Models (AAMs) [11] for time-series facial expressions at the beginning, a feature vector sequence was obtained for each part of the face. Then, they acquired expressive notation of involuntary and spontaneous facial expressions based on providing an automatic phrase of the mode from the obtained feature vectors. The experimentally obtained results show that by particularly addressing timing structures of the two expressive notations that were obtained, Hirayama et al. analyzed how two facial expressions differ. In the analytical results, a difference was found in the timing of movement of the muscles between lifting the cheek and moving the mouth for two facial expressions. Consequently, for describing the timing structure of facial expressions, the time resolution of the model and the image sequence are set high using expressive notation. However, because the spatial resolution of the model representing facial expressions is low, analysis of differences of the expressive intensities representing the intermediate facial expression have not yielded satisfactory results.

Otsuka et al. [12] proposed a method to extract six individual basic expressions described by Ekman et al. [13]. Modeling the movement of the facial expressions by HMMs, which carry out the state transition corresponding to the motion of different facial muscles, i.e., relaxation, contraction, stationary, and elongation, Otsuka et al. sought to recognize the facial expressions by analyzing the motion vectors of their surroundings, noticing that AUs of Facial Action Coding System (FACS) are distributed around the eyes and mouth. In their method, they first obtained the motion vectors around the eyes and mouth using the gradient method [14] from the facial expression image sequences, e.g., facial expressions of two kinds for 20 subjects. Next, by performing two-dimensional Fourier transform in a matrix component of the image, they

 ISBN: 978-1-61208-356-8

acquired a time series of 15-dimensional feature vectors. As the input time series of the feature vectors, Otsuka et al. extracted individual facial expressions by application of HMMs of left-to-right type. In this case, the experimenters confirmed the determination of true or false facial expressions. In a section of actual facial expressions, they treated the corresponding facial expression that had been extracted as a correct answer. In contrast, they treated the following two cases as incorrect answers: when no facial expression was extracted; when different facial expressions were extracted at once. An extraction rate of 90% was achieved in their experimentally obtained results: 40 facial expressions were extracted in the 20 subjects. Then, 36 facial expressions were accurately extracted in them. However, it is not always the precise period because being extracted represents the start and end points of facial expressions. The correctness checker is treated as a correct answer when the corresponding facial expression is expressed within the period.

According to the most recent study of the emotion–expression relationship based on evidence from laboratory experiments [15], high coherence has been found in several studies between amusement and smiling; low to moderate coherence between other positive emotions and smiling. Additionally, insufficient emotion intensity and inhibition of facial expressions could not account for the observed dissociations between emotion and facial expression. Furthermore, as a statistical indice of the coherence between emotion and facial expression, R. Reisenzein et al. reported that the most informative indice was "the average intra-individual correlation between emotion and expression". In this study, we actively do challenge to elucidate the correlations between the expressive process of individual-specific facial expressions and psychological states, particularly focusing on the correlations between pleasant-unpleasant stimulus and smiling process of intentional facial expressions.

III. FRAMEWORK OF EXPOSED RHYTHMS AND TEMPOS

A. *Facial Expression Levels*

As an index for quantifying the individual facial expression spaces, we proposed the framework of expression levels (ELs) [3]. The ELs include both features of the pleasure and arousal dimensions based on the arrangement of facial expressions on Russell's circumplex model [16]. Specifically, we extract the dynamics of topological changes of facial expressions of facial components such as the eyes, eyebrows, and mouth. Here, topological changes show the structure defining the connection form of the elements in the set [2]. The ELs obtained in this study are sorted categories according to their topological changes in intensity from expressions that are regarded as neutral facial expressions. As discussed above, the ELs in this study include both features of the pleasure and arousal dimensions. In Russell's circumplex model, all emotions are constellated on a two-dimensional space: the pleasure dimension of pleasure–displeasure and arousal dimension of arousal–sleepiness. In the intentional facial expressions covered in this study, directly handling the facial expressions for the influence of pleasure dimension is difficult. Therefore, as a method of measuring transitory stress response, we conduct an evaluation using the salivary amylase test. Therefore, as a method of measuring transitory stress response, we conduct an evaluation using the salivary amylase test through the task of watching emotion-evoking videos caused a pleasant-unpleasant state. Focusing on the values of salivary amylase activity between before and after watching videos, we can effectively perform stress measurements by the salivary amylase test to assess the stress state transiently. Consequently, we target the intentional facial expressions under stimulating states of pleasant and unpleasant.

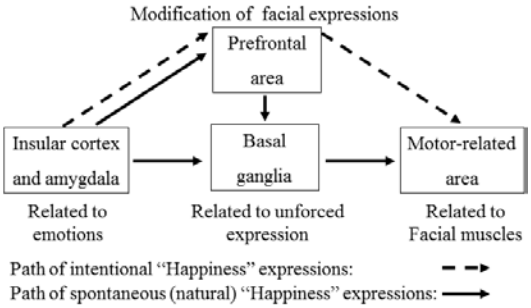


Figure 1. Expression paths based intentional and spontaneous facial expressions.

B. *Definition of Exposed Rhythms and Tempos*

Blair [17] has reported that, for facial expressions, four brain domains are mutually related: (1) parts producing feelings (insular cortex and amygdala), (2) parts forming facial expressions involuntarily (basal ganglia), (3) parts embellishing facial expressions according to the surrounding circumstances (prefrontal area), and (4) motor-related areas actually moving mimic muscles. Yamaguchi [18] reported that the brain memorizes experiences in a rhythm: according to specific brain waves, nerve cells work cooperatively, and experiences are memorized. In perceptual recognition, it is explained that nerve cells function simultaneously according to the gamma waves, which are brain waves having quick rhythms. From the results of these studies, we infer that the rhythms of nerve cells participate in the expressional process of facial expressions. As presented in Figure 1, in cases where facial expressions are embellished intentionally or spontaneously, time-sequential differences exist based on the route through

 ISBN: 978-1-61208-356-8

which facial expressions are revealed. The basis of our hypothesis is as follows. According to specific brain waves of four brain area, nerve cells of each brain area are used to work cooperatively, in the case of the repetition process of facial expressions under a pleasant-unpleasant stimulus particularly. Mimic muscles is activated by coordination of nerve cells with different speed, a unique expression is exposed through the individual path of each facial expression.

In this study, using temporal variation of ELs, we intend to visualize rhythms and tempos of facial expressions that humans create. We defined one rhythm as a tempo that is repeated several times. One tempo indicates the period during which facial expressions are transformed from a neutral state to the next neutral state. Facial expressions exposed intentionally by humans form an individual space based on dynamic diversity and static diversity of the human face [19]. Facial expression dynamics can be regarded as "topological changes in time-sequential facial expression patterns that facial muscles create." Static diversity is individual diversity that is configured by the facial componential position, size, and location, consisting of eyes, nose, mouth, and ears. In contrast, dynamic diversity represents that human can move facial muscles to express internal emotions unconsciously and sequentially or to express emotions as a message. After organizing and visualizing topological changes of face patterns by ELs, we attempt to use the framework of rhythms and tempos with expressions to express ambiguities and complexities of facial expressions attributable to a psychological state.

IV. PROPOSED METHOD

Facial expression processes differ among individuals. Therefore, Akamatsu [19] described the adaptive learning mechanisms necessary for modification according to individual characteristic features of facial expressions. In this study, our target is intentional facial expressions. We use Self-Organizing Maps (SOMs) [20] to extract topological changes of facial expressions and for normalization with compression in the direction of the temporal axis. After classification by SOMs, facial images are integrated using Fuzzy ART [21], which is an adaptive learning algorithm with stability and plasticity. In fact, SOMs perform unsupervised classification input data into a mapping space that is defined preliminarily. In contrast, Fuzzy ART performs unsupervised classification at a constant granularity that is controlled by the vigilance parameter. Therefore, using SOMs and Fuzzy ART, time-series datasets showing changes over a long term are classified with a certain standard. Figure 2 presents an overview of the procedures used for our proposed method. In the following, we describe extraction of time-sequential changes of ELs, and also explain detection of expressive

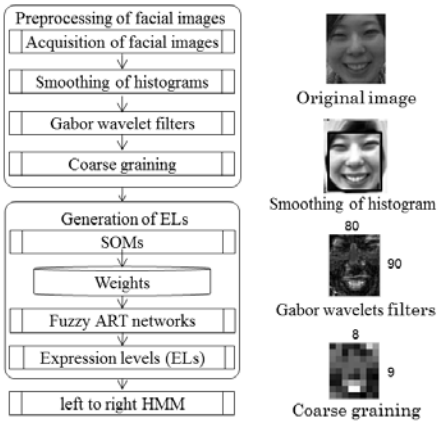


Figure 2. Overview of the procedures used for our proposed method.

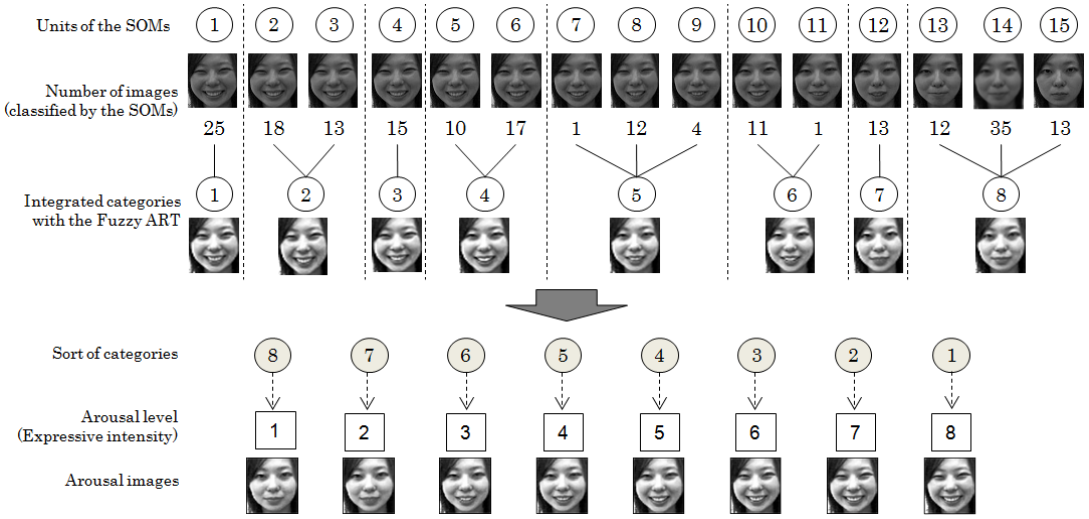


Figure 3. Procedure details for acquiring a time-series variation of ELs.

tempos by HMMs.

A. Acquisition of Time-series Variation of ELs

We set the Region of Interest (ROI) to 90 × 80 pixels, including the eyebrows, which all contribute to the impression of a whole face as facial feature components. With preprocessing, brightness values are normalized for time-series images of facial expressions. The influence of brightness values attributable to illumination conditions is thereby reduced. Moreover, smoothing the histogram is useful to adjust contrast and clarify the images. In addition, using the orientation selectivity of Gabor Wavelets filtering as a feature representation method, the facial parts characterizing the dynamics of facial expressions are emphasized, such as eyes, eyebrows, mouth, and nose. By down-sampling (i.e., 10 × 10 pixels) time-series facial expressions converted with Gabor Wavelets filtering [22], the effects of a slight positional deviation when taking facial images are minimized. Then data size compression is conducted.

Figure 3 presents details of procedures for acquiring a time-series variation of ELs. First, we use SOMs to learn the time-series images of facial expressions with down-sampling. The face images that show topological changes of facial expressions that are similar are classified into 15 mapping units of SOMs. Next, similar units (i.e., Euclidean distances of the weight vectors are close) among 15 mapping units of SOMs are integrated into the same category by Fuzzy ART. By sorting the facial expression categories integrated by Fuzzy ART from neutral facial expression to the maximum of facial expression, we obtain ELs labeled as expressive intensities of facial expressions quantitatively. The sorting procedure of integrated categories is based on the two-dimensional correlation coefficient of the average image of the facial expression images classified into each category. Finally, we conduct corresponding ELs with each frame of the facial images to produce time-series variations of ELs.

B. Extraction of Expressive Tempos by HMMs

As a method of recognizing words by estimating phonemes from acoustic signals, HMMs were first used in the speech recognition field. Takeda et al. [23] performed an automatic accompaniment and score tracking of MIDI music using HMMs. Actually, HMMs have been established as a technique for extracting an occurrence pattern from time-series datasets and classifying it as a category. Datasets used for this study are directed to time-series facial images, an expressive tempo consists of occurrence pattern of ELs. Therefore, we use HMMs to extract expressive tempos. HMMs are simple Markov models with multiple nodes, defined by transition probabilities between mutual nodes and output probabilities of multiple symbols from each node. By preparing HMMs to extract a target, each HMM is trained in the symbol sequence of each training dataset for

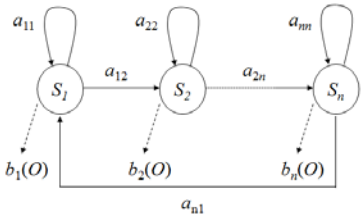


Figure 4. Configuration of HMMs used for this study (Type of Left to Right).

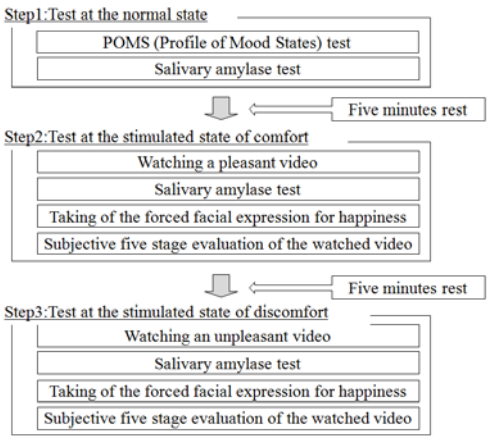


Figure 5. Details of experimental protocols.

targets. Training of HMMs is useful to estimate two parameters of symbol output probabilities and state transition probabilities that generate a high probability of training symbol sequence. Additionally, using Baum–Welch algorithm [24], training is repeated until the parameters converge i.e., the change in the output likelihood is sufficiently small. The configuration of HMMs used for this study is a type of Left to Right, as shown in Figure 4, we set the internal state of nodes to S_1, S_2, · · ·, S_n from left to right. Here, S_1 is the initial state of facial expressions (neutral facial expression), S_2 · · ·S_{n-1} are the intermediate states, and S_n is designated as the final state (maximum value of ELs). To obtain the updated values of state probability of S_i (i = 1, · · ·, n), we define the probability of following equations. State transition probabilities (a_{ij}) mean the transition probability from state S_i to state S_j, only the self-transition and transition to the right state in Left to Right HMMs are permitted. Therefore, the following constraints are satisfied.

$$a_{ij} = 0 \qquad (j < i) \tag{1}$$

$$0 <= a_{ij} <= 1 \qquad (j >= i) \tag{2}$$

$$\sum a_{ij} = 1 \tag{3}$$

Symbol output probabilities bi (O) denote the probability density distribution for outputting a symbol sequence O in state S_i, we use a discrete distribution of allocating

probabilities to discrete symbols that are commonly used in the field of speech recognition.

V. DATASETS

In this study, we constructed an original and long-term dataset for the specific facial expressions of one subject. Figure 5 presents details of experimental protocols. One experiment comprises three steps, i.e., step 1 is under a normal state, step 2 is in watching pleasant video, and step 3 is in watching unpleasant video. As shown in Figure 6, we gave subjects the task of watching emotion-evoking videos caused a pleasant–unpleasant state, and performed stress measurements by salivary amylase tests to assess the stress state transiently. In addition, the watching time is about 3 min for each emotion-evoking video, we prepared unpleasant videos (i.e., implant surgery and cruel videos) and the pleasant videos (i.e., comic videos of three types). The subjective assessment of five stages was also conducted at watching videos. For all subjects, we fully explained the experiment contents in advance based on the research ethics policy of our university, and also obtained the consent of experiment participants in voluntary writing of subjects. Moreover, from all subjects, we received agreement to publish face images as part of their experimental participation.

A. Facial Expression Images

Open datasets of facial expression images are open to the public through the internet from universities and research institutes. However, the specifications vary among datasets because of imaging with various conditions. As static facial images, the dataset presented by Ekman and Friesen [13] is a popular dataset comprising collected various facial expressions used for visual stimulation in psychological examinations of facial expression cognition. As dynamic facial images, the Cohn–Kanade dataset [6] and Ekman–Hager dataset [25] are widely used, especially in experimental applications. In recent years, the MMI Facial Expression Database presented by Pantic et al. [8] and the CK+ dataset [7] have become a widely used open dataset containing both static and dynamic facial images. These datasets contain a sufficient number of subjects as horizontal datasets. However, facial images are taken only once for each subject. No dataset exists in which the same subject has been traced over a long term. Therefore, we created original and longitudinal datasets that include collections of the specific facial expression of the same subject during a long term.

Six basic facial expressions proposed by Ekman et al. [13] are "happiness", "anger", "sadness", "disgust", "fear", and "surprise". Among the six basic facial expressions, we specifically examined the facial expression of "happiness", which is believed to be most likely exposed spontaneously. As the target facial expression of "happiness" under stimulating states of pleasant and unpleasant, we acquired the facial expressions of 20 subjects. As a method of stimulation, we pre-selected emotion-evoking videos that elicit emotions that are pleasant or unpleasant, with all subjects expressing the facial expression of "happiness" immediately after watching them. Subjects were 10 men (Subject J was 20 years old; Subjects B, G, H, and I were 21; Subjects A, E, and F were 22; Subjects C and D were 23) and 10 women (Subjects K, M, O, and P were 20 years old; Subjects L, Q, R, S, and T were 21; Subject N was 23), all of whom were university students. The imaging period was three weeks at one-week intervals for all subjects. The imaging environment for facial expressions was an imaging space partitioned by a curtain in the corner of the room. We took frontal facial images with conditions including the head of the subject in each image. In advance, we instructed each subject to expose the facial expression without any head movement. Consequently, imaging the face region to fit within the scope has been possible. However, with respect to extremely small changes caused by body motion, we used template-matching methods to trace the face region by setting the initial template to include facial parts. By consideration of the application deployment and ease of imaging in future studies, we used commercially available USB cameras (QcamOrbit; Logicool Inc. [26]). When taking images of each facial expression, the same expression was repeated three times based on the neutral facial expression during the image-taking time of 20 s. We previously instructed all subjects to express an emotion three times at their own timing according to a guideline for 20 s. One dataset consisted of 200 frames with the sampling rate of 10 frames per second.

B. Stress Measurement Method

Because types of psychological stress are regarded as affecting facial expressions, we assessed transient stress and chronic stress. Chronic stress is that which humans have on a daily basis, whereas transient stress is that caused by a temporary stimulus. To assess transient stress stimulus to the subjects in this study, we applied the salivary amylase test, which is one method of measuring transient stress reactions. As a biological reaction, salivary amylase activity is detected as a low value if one is in a pleasant state. In contrast, the value is high if one is in an unpleasant state. As stress reactions when subjected to external transient stimulus, Yamaguchi et al. [27] confirmed that salivary amylase activity is an effective means of stress evaluation. For this study, using the emotion-evoking videos as an external transient stimulus, we used the salivary amylase test method to measure stress reactions immediately after participants watched the videos.

 ISBN: 978-1-61208-356-8

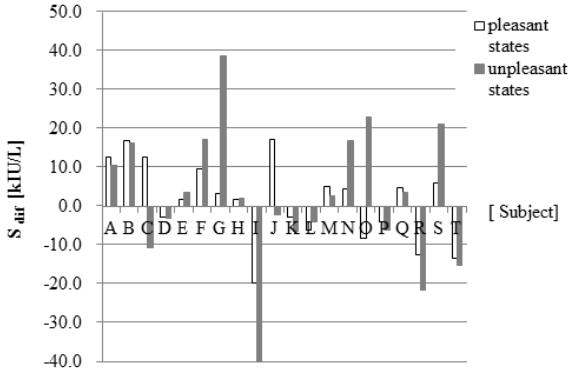


Figure 6. Results of S_{dif} obtained for target to the 20 subjects of A-T.

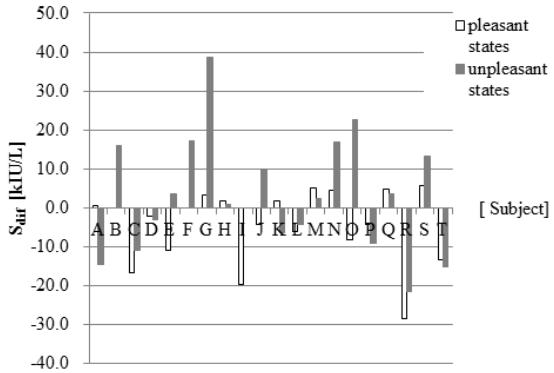


Figure 7. Results of S_{dif} addressed only the score of 4 and 5 with subjective evaluations.

VI. EXPERIMENT

We verified the validity of emotion-evoking videos, which give a pleasant–unpleasant stimulus. Next, we optimized the number of states of HMMs by extracting expressive tempos from facial expressions. Subsequently, using the HMMs with an optimized number of states, we verified the accuracy of the extracted expressive tempo obtained from a time-series change of ELs. Finally, we analyzed the transient stress stimulus of pleasant–unpleasant effects on the expressive rhythm of facial expressions.

A. Effectiveness of Pleasant–unpleasant Stimulus

Using the salivary amylase test, we examined the validity of emotion-evoking factor in watching the video used as a pleasant–unpleasant stimulus. The following were shown for salivary amylase activity. The value of salivary amylase activity is reduced if in a pleasant state. In contrast, its value is increased if one is in unpleasant circumstances [27]. Accordingly, letting S_{normal} be the value of salivary amylase activity at normal state, and letting S_{stimu} be the value of salivary amylase activity after watching the video, then the difference of salivary amylase activity between the normal state and after watching video (S_{dif}) is defined by the following equation.

$$S_{dif} = S_{stimu} - S_{normal} \quad (4)$$

$S_{dif} < 0$ (i.e., after watching pleasant videos)

$S_{dif} > 0$ (i.e., after watching unpleasant videos)

Figure 6 presents results of S_{dif} obtained for target to the 20 subjects of A–T. In this case, the perception for the pleasant–unpleasant videos differs slightly among subjects, so this fact might cause the results of salivary amylase activity of C and B differ with previous studies [27]. Therefore, we decided to calculate the salivary amylase activity only for data for which subjective evaluation of the subject is high. The subjective evaluation receives a score of 1–5, score 1 (i.e., not at all), score 5 (i.e., strong) at watching each emotional video. Figure 7 presents results of salivary amylase activity in the case of particularly addressing only the score of 4 and 5 because we consider that the emotional video is effectively working as a pleasant–unpleasant stimulus. Based on this result, the average of all S_{dif} indicates -2 [kIU/l] at a pleasant state, 5 [kIU/l] at an unpleasant state. Therefore, results show that the emotion-evoking video functioned as a pleasant–unpleasant stimulus.

B. Examination of HMM Parameters

Otsuka et al. [12] pointed out that the process of facial expressions was made up with state transitions such as "neutral state" → "expression state" →"neutral state". In this case, the operation of facial muscles was to be the acts of "relaxation" → "contraction" → "rest" → "extension" → "relaxation". In the method explained by Otsuka et al., under conditions in which the state of facial muscles and the state of HMMs are associated with initial values, they modeled the state transitions of facial muscles by setting the number of states of HMMs to five [12]. However, by varying the initial state transition probability and number of states of the HMMs, our experiments were conducted to ascertain the optimum value of the highest extraction rate shown in equation (5). Therefore, it is possible to obtain parameters (i.e., the initial state transition probability and number of states of the HMMs) that represent the best movement of facial muscles under conditions of transient stress stimulus.

As the accuracy judgment of extraction with HMMs, we set as Ground Truth (GT) the average value of the frames for which three evaluators judged that the transition state had returned to a neutral state by their visual observation of the videos showing facial expressions. The extraction rate of accuracy judgment is defined by equation (5).

$$x_1, x_2, x_3 = \begin{cases} 1, \mathrm{E} \in \mathrm{R} \pm 5 \\ 0, \neg(\mathrm{E} \in \mathrm{R} \pm 5) \end{cases}$$

$$A = \frac{x_1, x_2, x_3}{C} \times 100[\%] \quad (5)$$

In that equation, *A* represents the extraction rate, *C* denotes the number of facial expressions, *E* represents the final frame of the facial expressions extracted with HMMs, and *R* denotes the frame indicating the end of the facial expressions obtained as the GT.

For this study, we performed experiments by obtaining the number of states to represent the movement of facial muscles optimally in a stress stimulus. In the pleasant–unpleasant state, we compared the extraction rate by varying the transition probability *b* to the next state, the self-transition probability *a*, and a number of states of the HMMs. Figure 8 presents the results. In the experimentally obtained result, the average extraction rate is the largest with setting the number of states to three. The average extraction rate is reduced later peaked at the state number of 3. Furthermore, the average extraction rate becomes a maximum under conditions of self-transition probability *a* of 0.70, and state transition probability *b* of 0.30. Based on consideration of the results described above, the parameters of the HMMs in this study were determined as follows. The number of states is 3, the self-transition probability *a* is 0.70, and the state transition probability *b* is 0.30.

C. Extraction Results of Expressive Tempos

As the extracted results of expressive tempos by application of HMMs, Figure 9 depicts the expressive tempos of six cases of subjects A, C, J, K, Q, and S. The top of each figure shows the time-series change of ELs. The bottom of each figure shows the transitional state of HMMs. Additionally, we marked the dashed vertical lines as GT. The GT indicates the average value of the frames, in which three evaluators judged that the facial expression had been completed by their visual observation for the original image. In consideration of variation among three evaluators, we

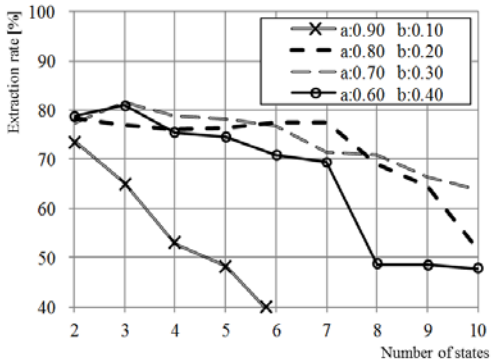
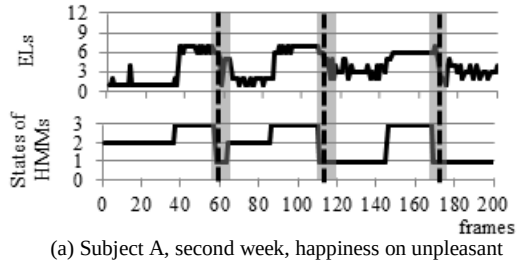
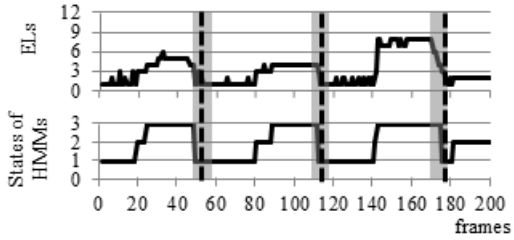


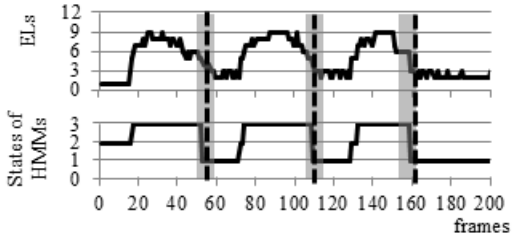
Figure 8. Extraction rates by varying a number of states of HMMs.



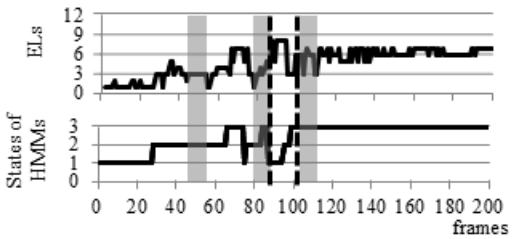
(a) Subject A, second week, happiness on unpleasant



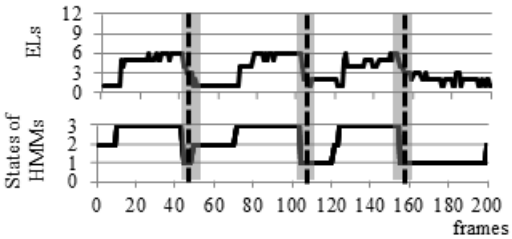
(b) Subject C, third week, happiness on unpleasant



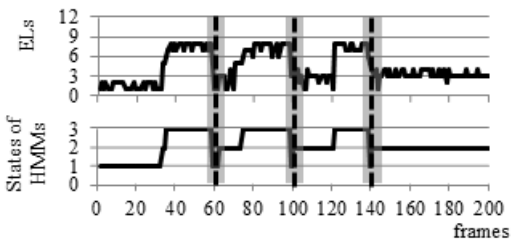
(c) Subject K, second week, happiness on pleasant



(d) Subject J, first week, happiness on pleasant



(e) Subject Q, first week, happiness on pleasant



(f) Subject S, first week, happiness on unpleasant

Figure 9. Extracted results of expressive tempos for six subjects.

 ISBN: 978-1-61208-356-8

presented a shaded gray pattern as the period of extraction, indicating a range of ± 5 frames with respect to each GT frame.

For subjects A, C, K, Q, and S, the extraction rates are 100% because all frames extracted by HMMs were included in the extraction range. In subject J, the extracted frames by HMMs are 60, 76, and 88, whereas the frames of GT are 40, 71, and 98. Therefore, in this example, only the second tempo was extracted successfully. Turning to the time-series change of ELs in the top of figure, extraction results of HMMs do not correspond to the timings of facial expressions. A major cause of that lack of correspondence is that evaluators have difficulty dividing the periods of facial expressions by visual observation because expressive levels of facial expressions appearing on the original image are small. For this study, we used view-based feature representation of facial expression datasets. Given difficulty in identifying the periods of facial expressions by human visual observation, we believe that automatic extractions of expressive tempos generally become difficult. Therefore, when acquiring facial expression datasets, we must ensure an instruction for each subject to expose the maximum ELs possible.

Subsequently, targeting the facial expression datasets of three weeks for subjects A–T (i.e., 20 cases), Figure 10 presents extraction rates of expressive tempos for each subject. Taking the average of the extraction rates in three weeks, the pleasant state was 81.1%. The unpleasant state was 77.8%. Even including a difficult case of identification of the facial expression period by visual inspection, such as Figure 10(d), the average extraction rate of 79.5% was obtained for all subjects.

D. Effects of Pleasant–unpleasant State on Expressive Rhythms

For subject G, Figure 11 presents the extraction result of expressive tempos and the time-series variation of ELs with "happiness" after watching pleasant videos. The three extracted tempos are as follows. The first tempo comprises 60 frames, the second tempo comprises 57, and the third tempo comprises 36. As described above, there are variations in the three expressive tempos which constitute one rhythm. Therefore, by calculating the average and standard deviation of number of frames constituting one tempo for all subjects A to T, we discuss the relation of expressive rhythms with a pleasant–unpleasant state.

Table I presents the standard deviation of tempos and average number of frames constituting one expressive tempo for all subjects of three weeks. Considering the average frames constituting one tempo in the pleasant–unpleasant state, the pleasant state is 49.1 [frames], the unpleasant state is 49.2 [frames]. Therefore, we conclude that the pleasant–unpleasant state does not affect the average number of frames that constitute one tempo. In contrast, particularly addressing the standard deviation of the number of frames constituting one tempo, the pleasant state is 8.4 [frames]; the unpleasant state is 6.1 [frames]. Comparison of the pleasant state and unpleasant state showed variation in the unpleasant state in the frames constituting one tempo. As a tendency among all subjects by transient stress stimulus watching unpleasant videos, we demonstrated quantitatively that fluctuations occurred in expressive tempos that were components of the expressive

TABLE I. STANDARD DEVIATION OF TEMPOS AND AVERAGE NUMBER OF FRAMES CONSTITUTING ONE TEMPO FOR ALL SUBJECTS

	Pleasant states	Unpleasant states
Average of frames	49.1	49.2
Standard deviation	6.1	8.4

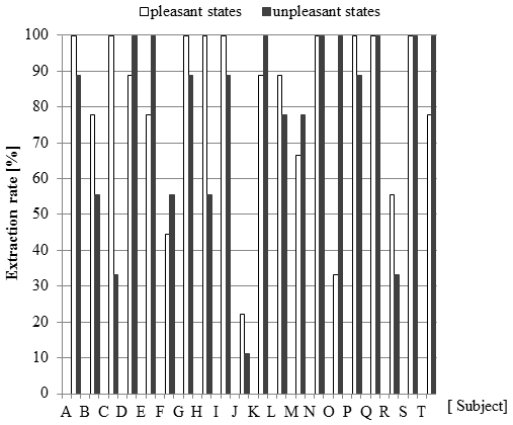


Figure 10. Extraction rates of expressive tempos for each subject.

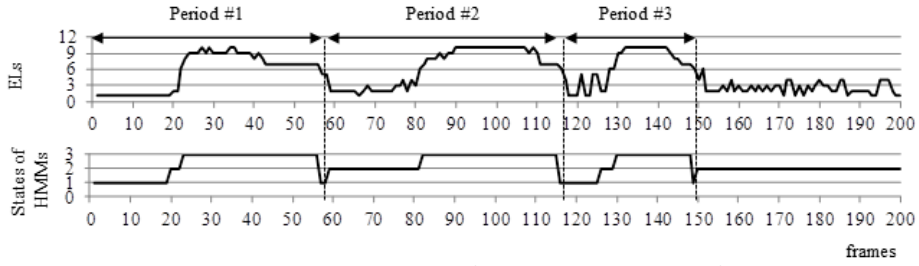


Figure 11. Expressive tempos and the time-series variation of the ELs with “happiness” after watching pleasant videos.

rhythm. The results described above reveal one indicator estimating the psychological state of humans. We conclude that the analysis of expressive tempos and rhythms is valid, with emphasis on repeated operations of intentional facial expression with "happiness".

VII. CONCLUSION AND FUTURE WORK

In this study, using the framework of expressive tempos and rhythms in facial expressions, we examined the relation between the psychological state (i.e., pleasant or unpleasant) and the time-series variation of ELs with exposure of intentional facial expressions. Acquiring image datasets of facial expressions under the states of pleasant–unpleasant stimulus for 20 subjects, we extracted expressive tempos of each subject. Consequently, taking the average of the extraction rates, the pleasant state was 81.1%, and the unpleasant state was 77.8%. By taking the effects of pleasant–unpleasant stimulus on the expressive tempos, particularly addressing the variation of number of frames constituting one tempo, the variation in unpleasant stimulus became greater than that in pleasant stimulus. The results presented above demonstrate that analysis using expressive tempos and rhythms is valid to indicate the psychological state. Moreover, by quantifying fluctuations of expressive tempos and rhythms, we can ascertain differences of the expressive path between intentional and spontaneous facial expressions.

ACKNOWLEDGMENTS

The authors thank the 20 students at our university who participated as subjects by letting us take facial images over such a long period. This work was supported by the Japan Society for the Promotion of Science (JSPS) KAKENHI Grant Number 25330325 and the Cosmetology Research Foundation.

REFERENCES

[1] N. Nobutani and Y. Nakatani, "Talk support system by rhythmical sound based on personal tempo," Information Processing Society of Japan, The 71st National Convention, pp. 4.227-4.228, Mar. 2009.

[2] S. Ohishi and M. Oda, "The personal tempo's effect on dialogue smoothness – from an index of switching pause –," The Institute of Electronics, Information, and Communication Engineers, Technical Report, pp. 31-36, 2005.

[3] H. Madokoro, K. Sato, and S. Kadowaki, "Facial expression spatial charts for representing time-series changes of facial expressions," Japan Society for Fuzzy Theory, vol. 23, no. 2, pp. 157-169, 2011.

[4] T. Sakaguchi, J. Ohya, and F. Kishino, "Facial Expression Recognition from Image Sequence Using Hidden Markov Model," The Institute of Image Information and Television Engineers, vol. 49, no. 8, pp. 1060-1067, 1995.

[5] H. Kumano, "Stress evaluation," http://hikumano.umin.ac.jp/StressAssess.pdf [retrieved: July, 2014]

[6] T. Kanade, J. F. Cohn, and Y. L. Tian, "Comprehensive database for facial expression analysis," Proc. of the 4th IEEE Int. Conf. on Automatic Face and Gesture Recognition, pp. 46-53, 2000.

[7] P. Lucey et al., "The Extended Cohn-Kanade Dataset (CK+): A complete expression dataset for action unit and emotion-specified expression," Proc. of the 3rd Int. Workshop on CVPR for Human Communicative Behavior Analysis, pp. 94-101, 2010.

[8] M. Pantic, M.F. Valstar, R. Rademaker, and L. Maat, "Web-based database for facial expression analysis," Proc. IEEE Int'l. Conf. Multimedia and Expo, Amsterdam, The Netherlands, Jul. 2005. doi: 10.1109/ICME.2005.15214.

[9] S. Das and K. Yamada, "Evaluating instantaneous psychological stress from emotional composition of a facial expression, " Journal of Advanced Computational Intelligence and Intelligent Informatics, vol. 17, no. 4, pp. 480-492, 2013.

[10] T. Hirayama, H. Kawashima, M. Nishiyama, and T. Matsuyama, "Facial expression representation based on timing structures in faces," Human interface: the transaction of Human Interface Society, pp. 271-281, May 2007.

[11] T. F. Coots, G. J. Edwards, and C. J. Taylor, "Active Appearance Model: Proceedings of European Conference on Computer Vision," vol. 2, pp. 484-498, 1998.

[12] T. Otsuka and J. Ohya, "A study of spotting segments displaying facial expression from image sequences using HMM," The Institute of Electronics, Information, and Communication Engineers, Technical Report, pp. 17-24, Nov. 1997.

[13] P. Ekman and W. V. Friesen, "Unmasking the face: a guide to recognizing emotions from facial clues," Malor Books, 2003.

[14] B. K. P. Horn and B. B. Schunck, "Determining optical flow," Artificial Intelligence, vol. 17, pp. 185-203, 1981.

[15] R. Reisenzein, M. Studtmann, and G. Horstmann, "Coherence between emotion and facial expression: evidence from laboratory experiments, " Emotion Review, vol. 5,no. 1, pp. 16-23, 2013.

[16] J.A. Russell and M. Bullock, "Multidimensional scaling of emotional facial expressions: similarity from preschoolers to adults," Journal of Personality and Social Psychology, vol. 48, pp. 1290-1298, 1985.

[17] R.J.R. Blair, "Facial expressions, their communicatory functions and euro-cognitive substates," Philos. Trans. R. Soc. Lond., B358, pp. 561-572, 2003.

[18] Y. Yamaguchi, "Contextual information rhythmically processed in the brain," IEEJ Transactions on Electronics, Information and Systems C, vol. 128, no. 8, pp. 1068-1071, Aug. 2000.

[19] S. Akamatsu, "Recognition of facial expressions by human and computer [I]: facial expressions in communications and their automatic analysis by computer, " The Journal of the Institute of Electronics, Information, and Communication Engineers, vol. 85, no. 9, pp. 680-685, Sep. 2002.

[20] T. Kohonen, Self-organizing maps, Springer Series in Information Sciences, 1995.

[21] G. A. Carpenter, S. Grossberg, and D.B. Rosen, "Fuzzy ART: fast stable learning and categorization of analog patterns by an adaptive resonance system," Neural Networks, vol. 4, pp. 759-771, 1991.

[22] M. Haghighat, S. Zonouz, M. Abdel-Mottaleb, "Identification Using Encrypted Biometrics, " Computer Analysis of Images and Patterns, Springer Berlin Heidelberg, pp. 440-448, 2013.

[23] H. Takeda, T. Nishimoto, and S. Sagayama, "Automatic accompaniment system of MIDI performance using HMM-based score following," Information Processing Society of Japan, SIG Technical Report, pp. 109-116, Aug. 2006.

[24] The Institute of Electronics, Information and Communication Engineers Edition, "Speech recognition by probabilistic model," Corona Publishing Co. Ltd., ISBN: 978-4-88552-072-3, pp. 29-66, 1988.

[25] M. Bartlett, J. Hager, P. Ekman, and T. Sejnowski, "Measuring facial expressions by computer image analysis," Psychophysiology, vol. 36. pp. 253-264, 1999.

[26] QcamOrbit; Logicool Inc., http://www.logicool.co.jp/ja-jp/webcam-communications/webcams [retrieved: July, 2014]

[27] M. Yamaguchi, T. Kanamori, M. Kanemaru, Y. Mizuno, and H. Yoshida, "Correlation of stress and salivary amylase activity," Japanese Journal of Medical Electronics and Biological Engineering: JJME vol. 39, no. 3, pp. 46-51, Sep. 2001.

 ISBN: 978-1-61208-356-8