IEICE The Institute of Electronics, Information and Communication Engineers

IEICE Proceeding Series

Estimating Motion Segments Using CNV-like Variation

Mika Sunagawa, Masataka Niwa, Hiroyuki Iizuka, Hideyuki Ando, Taro Maeda

Vol. 1 pp. 735-738
Publication Date: 2014/03/17
Online ISSN: 2188-5079

Estimating Motion Segments Using CNV-like Variation

Mika Sunagawa, Masataka Niwa, Hiroyuki Iizuka, Hideyuki Ando and Taro Maeda

Graduate School of Information Science and Technology, Osaka University
2-1 Yamadaoka, Suita, Osaka, JAPAN
+81-6-6879-7830
{sunagawa.mika, niwa, iizuka, hide, t_maeda}@ist.osaka-u.ac.jp

Abstract– This study aimed at investigating if CNV-like variations are elicited in Electroencephalograms (EEG) when people remember human continuous motion, and if it is possible to segment the continuous motion automatically by EEG. It is known that there are some important postures in continuous human motion, and the segments divided by the significant postures attract human attention. Since the situation is similar to Event-related brain potential (ERP) experiments, we expected that brain wave variations like Contingent Negative Variation (CNV) would also be elicited during motion remembering and it could be possible to detect the segmentation automatically from EEG. In this study, we used template matching and found that negative variations tend to precede segments divided by manually selected significant postures.

1. Introduction

It has been proposed that a user controls a robot intuitively using a low DOF (degree of freedom) controller to manipulate a high DOF robot by matching segmented human behavioral intention and segmented robot motion [1]. To realize this architecture, the segments of human behavioral intention and robot motion should be consistent and it becomes important to adjust the segments of the robot motion corresponding to the human behavioral intention since the behavioral intention is not adjustable. In our previous study, the segments of the robot motion were arbitrarily determined by an experiment designer to correspond with the human intention. However, it is not easy to estimate the segments and the breakpoints of human behavioral intention.

One of the ways to extract segments of behavioral intention is shown by Aso et al. [2] They asked participants to select significant frames from pre-recorded task motion sequences necessary for communicating the sequences so that another person might be able to reproduce the sequences from the selected frames. The selected frames of significant still images are regarded as the breakpoints of segments of behavior intention. Their result shows that the selected still images are enough to communicate the behavioral intentions and they can correctly reproduce the continuous motion after even watching only the still images. Although the method makes it possible to estimate the segments and breakpoints of human behavior intentions, it needs participants to select all motion segments one by one from motions, after which the robot motion segments can be designed manually.

Therefore, we investigate how segments and breakpoints of behavioral intentions can be extracted automatically. The present paper shows a method using brain activity. Since it is known that our attention would elicit Event-Related Potentials (ERP), it is expected that attention to the significant still images in the continuous motions would elicit ERP during motion remembering. By detecting the ERP, the segments of the behavioral intentions are estimated automatically.

2. Contingent Negative Variation

There are several kinds of ERP but we focus on the Contingent Negative Variation (CNV). CNV is elicited when people are concentrating on the interval between warning (S1) and imperative (S2) stimuli [3]. Those stimuli can be visual or auditory. After the S1, subjects start concentrating on the task to which they have been asked to react (e.g. pushing a button) soon after S2 is given. Figure 1 shows the waveform of CNV that Walter reported. During concentration, the brain wave develops negatively between S1 and S2, and it is known that the more a person concentrates, the larger is the CNV elicited. As another feature of CNV, the brain wave suddenly shifts to positive after the S2 stimuli. In the conventional studies, the CNV is well observed by signal averaging.

Since it is expected that significant still images in the continuous motions play the roles of S1 and S2 stimuli, CNV is used as cues of behavioural intentions, such that the segmentation of behavioral intentions can be captured with EEG automatically, whereas they are manually selected in Aso's experiments. We expected that negative variations like CNV would be elicited by significant still images in continuous motions during motion remembering. In our experiments this hypothesis is tested.

3. Task and Method

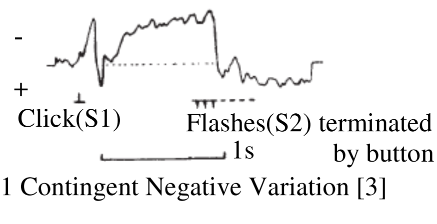

Fig.1 Contingent Negative Variation [3]

To investigate whether the CNV-like variation is elicited when participants are watching human continuous motion and whether there is a correlation between the variations and breakpoints of segments of behavioral intentions, we set up the following three tasks: letter-remembering (LR), motion-remembering (MR) and still-image-selecting (SIS) tasks. Four participants (4 males, ages between 22 and 29) joined our experiments.

3.1. Letter-remembering Task (LR)

To detect CNV-like variations elicited by the segmentation process when observing continuous human behavior, we used template matching. In order to obtain the template of the brain waveform that represents the most typical brain waves related to segmentation, participants were given explicitly segmented visual stimuli, which were four alphabetic letters displayed successively every 1.5 s on a monitor, and they were required to remember the letters. To make the situation similar to motion remembering task, alphabet characters were displayed for 7 s. The schematic view of the experimental sequence is shown in Fig. 2. At the beginning of a trial, “0” was displayed for 0.5 s. After a start marker was displayed for 167 ms (5 frames in 30fps), alphabet characters were displayed every 1.5 s. Between displayed characters, countdown numbers from “4” to “1” were displayed for 333 ms each (10 frames). Next, a randomly selected letter was displayed for 167 ms (5 frames). After showing the last letter, “0” was displayed again until the end. Participants observed 15 kinds of videos which display different alphabet character sequences and remembered the letters. Participants wrote down the displayed four letters after each trial.

During the letter-remembering task, EEG was recorded. After the experiment, a template of the brain waveform was created by EEG signals. The letters are explicitly segmented and the segments would elicit the CNV-like variations. The method of making the template is described in data-processing section.

3.2. Motion-remembering Task (MR)

In the MR task, participants observed and remembered human motion and EEG was recorded during observation. Five different kinds of human continuous motion were displayed on a monitor. The video sequence is shown in Fig.3. Each motion sequence consists of four different types of basic karate motions such as punching and kick-

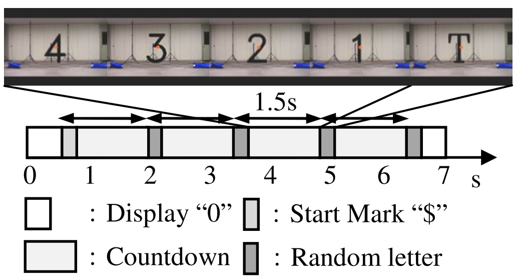

Fig.2 Components of alphabet video

ing. The length of the videos is 7 s same as the LR task. The actor keeps still for 0.5 s at the beginning. Each basic karate motion takes 1.5 s and the basic motions are connected smoothly. The actor keeps still for 0.5 s again at the end. Each video was displayed 15 times separated by 3 s rest. To enhance participants’ concentration, every 3 times they were required to reproduce the displayed motion by moving their own body as shown in Fig. 4.

In order to distinguish between concentration effects by segmentation and vision-related effects by moving objects, as a control task, participants also observed scrambled biological motion (SBM) before observing the human motion videos. Biological motion refers to the moving dots that reflect the motions of significant joints of human motion. Scrambled biological motion means that all recorded body points are randomly mapped to different positions. Scrambled biological motion shifts all dot positions of biological motion but the velocities of the dots are maintained. Figure 5 shows sample images of original human, biological and scrambled biological motion. Participants were asked to observe the scrambled biological motions in the same way as the MR task and not to analyze and concentrate on the moving dots. As shown in Fig.4, each video was displayed 15 times followed by 3 s rest, and the participants had an additional 10 s rest every three times. All recorded EEG data was evaluated by the template matching.

3.3. Still-image-selecting Task (SIS)

In the MR task, the participants were asked to only remember the continuous motion but the SIS task forced them to select still images out of the motions in the same

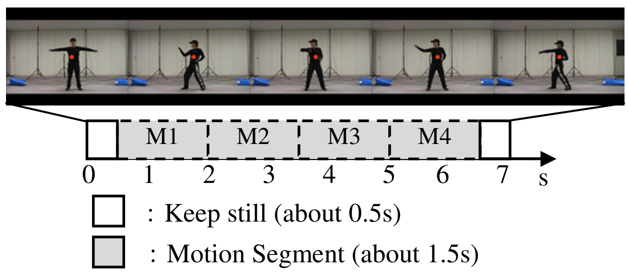

Fig.3 Components of motion video

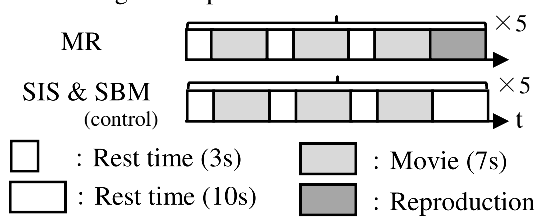

Fig.4 Timeline of motion-remembering, still-image-selecting and control task

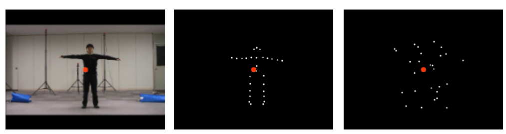

Fig.5 Human, biological and scrambled images

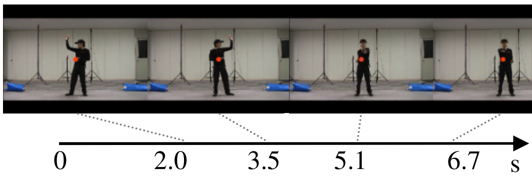

Fig.6 Sample of selected still images

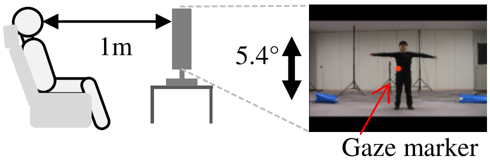

Fig.7 The experimental environment

way as Aso's experiments after watching videos. By concentrating on selecting the still images, CNV-like variations could be elicited more than or same as the MR task. The timings of selected still images are used to evaluate the correlations of detected timings of CNV-like variations in the MR task with the segments of behavioral intentions.

Following Aso's experiments, participants selected still images (postures) that are significant and necessary to communicate the continuous motion to other people, who have no idea of the motions, by showing only the selected still images. Figure 6 shows the sample of frames of selected still images. The selected frames are considered as breakpoints of motion segments of behavioral intention.

In the SIS task, motion videos were displayed 15 times in the same ways as the MR task as shown in Fig.4.

3.4. Electroencephalogram Acquisition

In all experiments, electroencephalograms (EEG) were recorded from Ag/AgCl pad electrodes placed at the midline central area (Cz; 50% of between nasion and inion) with the international 10-20 system. The earlobes were used for the reference electrodes and soles of the feet were earthed as body ground. As shown in Fig.7, participants were reclined in an easy chair. To detect eye-movement artifact such as eye-blink artifact, an electrooculogram (EOG) was also recorded using the Ag/AgCl electrode placed on the orbicularis oculi muscle above the left eye, and participants were instructed to gaze at a marker and not to move their eyes during video playback. The impedances of the electrodes were maintained below 20 kΩ. The recoded signals were amplified with a gain of 5000 and sampled by an AD converter (MP150ACE, BIOPAC) at a sampling frequency of 500 Hz. The filters used were the low-pass filter of 35 Hz and the high-pass filter of 0.05 Hz for EEG and 0.1 Hz for EOG.

3.5. Data Processing

Although the CNV is well observed by signal averaging in the conventional studies, the number of EEG waves we recorded is too few. To reduce background activities like alpha wave, we adopted a band-pass filter (0.5-6Hz). Invalid trials, in which eye-movements were detected (EOG>80μV), were removed in advance, the other trials' EEG waves are shifted to zero mean and filtered.

3.5.1. Make Template

All valid EEG waves for the LR task were averaged. Four time series of averaged EEG between 800 ms before and 400 ms after appearing alphabetic letters on the monitor were template candidates. The candidate which had the maximum negative area was adopted as the template. Figure 8 shows participant A's EEG waves and the template candidates. A solid line square shows A's adopted template.

3.5.2 Template Matching

The template created for each participant was used to find template-like EEGs by calculating correlation coefficients with the EEGs recorded in the MR and SIS tasks. The recorded data were grouped every five trials and averaged over valid trials to reduce single trial noise. Since there are 15 trials for the LR and MR tasks, we obtained three averaged data in each task and the correlation coefficient was calculated for each averaged EEG wave as follows. Let T(i) be a template and E(j) be an averaged EEG wave (i=0,1,…,N-1, j=0,1,…,M-1), correlation coefficient C(j) were calculated by the following equation:

$$C(j)=\frac{\sum_{i=0}^{N-1}((T(i)-\overline{T})(E(i+j-L)-\frac{1}{N}\sum_{k=0}^{N-1}E(k+j-L))}{\sqrt{\sum_{i=0}^{N-1}(T(i)-\overline{T})^2\times\sum_{i=0}^{N-1}(E(i+j-L)-\frac{1}{N}\sum_{k=0}^{N-1}E(k+j-L))^2}},\quad(1)$$

where L is the time lag between template onset and the time when letters appear. By this, the time of high correlation can be regarded as the time of estimated breakpoints of segments like significant frame of SIS task. Then, we defined the time of breakpoints of segments as the time when the time series of correlation coefficients satisfy the following two conditions: (1) It is over 0.5; (2) It is the maximal value during 600ms before and after. Condition

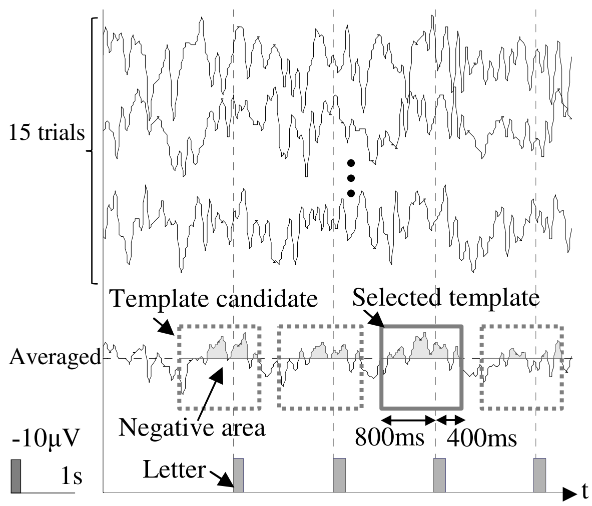

Fig.8 EEGs acquired from Cz and making a template

(2) does not allow overlapping over half of the template.

4. Result

Figure 9 shows the examples of averaged EEG data recorded in the MR and SIS task. Red lines indicate the correlation coefficients and small red filled areas represent the values over 0.5. The red arrow shows the estimated time of breakpoints of segments by EEG. The vertical dotted lines are the time that the participant explicitly selected the significant frames in the SIS task. To see the correlation between estimated and explicitly selected segmentations, histograms for each task for each participant A-D were made (Fig. 10). The zero in the x-axis is the time of explicitly selected significant frames (vertical dotted lines) and the values in the x-axis show the difference between time of the nearest selected significant frames and the estimated segmented points by EEG (red arrows). The estimated segmented points that were detected more than 400 ms after the significant frame were regarded as the EEG segmentation for the next selected frame. The histograms of Fig. 10 show the result over the five different videos under all tasks for each participant. The histogram has a high peak about 250-750 ms before the manually selected timings, and standard deviations (SD) of the delay in MR and SIS task tend to be lower than that of control. There is no large difference between the SIS and MR tasks.

There are two findings from our results. One is that CNV-like variation can be elicited when subjects remember human continuous motions because segmented information is stored in our brain to remember human motions and the timing of segmentations plays a role of S1 and S2 stimuli as they concentrate on those stimuli in the conventional CNV experiments. Another one is that negative shifts of the CNV-like variation were shifted to 250-750 ms before the timings manually selected as significant frames.

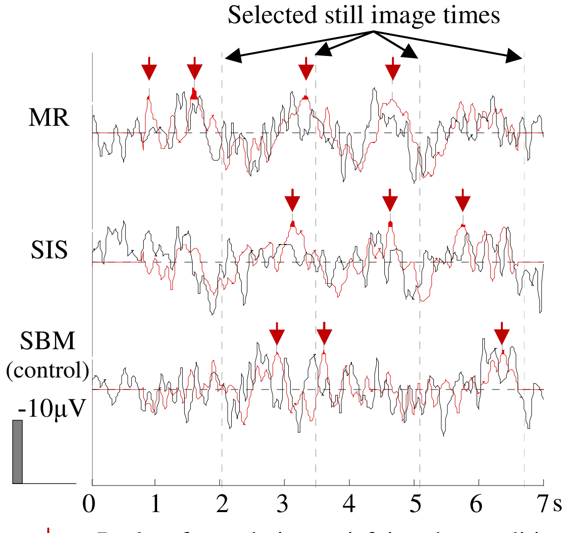

Fig.9 EEG and correlation coefficient of each task (participant A, movie 4, 1st part)

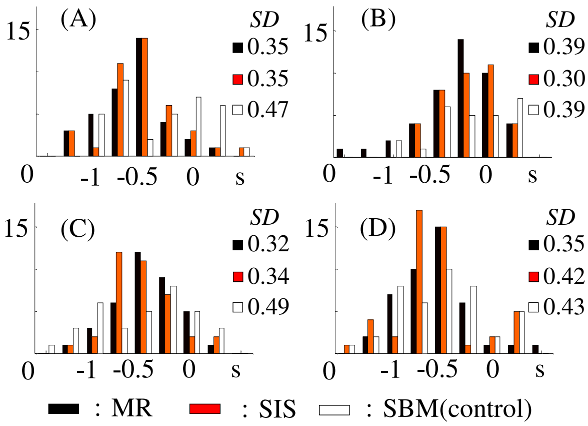

Fig.10 The histogram of the difference between time of the selected and estimated segmentation points

5. Discussion

Our result shows that CNV-like variations were elicited 250-750 ms earlier than expected. The possible reason is that participants could easily predict human motions since they watched each motion 15 times. On the other hand, in the LR task for making the template, participants could not predict the letters since they watched each sequence once. Therefore, there must be a time lag between two tasks in terms of concentration. To properly estimate the time as selected significant still frames from EEG, it is necessary to reveal the detailed time-correspondence between significant still frames and attention or to improve the way of making templates to avoid the prediction effect.

Acknowledgments

This work was partially supported by grant-in aid (No. 22240008) and also by a grant-in-aid from the Global COE program: Founding Ambient Information Society Infrastructure of the Ministry of Education, Culture, Sports, Science, and Technology, Japan.

References

[1] S. Okada, S. Sakaguchi, M. Niwa, H. Iizuka, H. Ando, and T. Maeda, "Investigation of Detection and Transmission of 'Tsumori', an Archetype of Behavioral Intention in Controlling a Humanoid Robot", *VRSJ the 14th Annual Conference, 2009*. (in Japanese)

[2] H. Aso, S. Okamoto, H. Kawasaki, H. Iizuka, H. Ando, and T. Maeda, "Communication and generation of motions by discretized / continuous visual images", *VRSJ the 13th Annual Conference, 2008*. (in Japanese)

[3] W. G. Walter, R. Cooper, V. J. Aldrodge, W. C. McCallum, and A. L. Winter, "Contingent negative variation: An electric sign of sensorimotor association and expectancy in the human brain," *Nature*, vol. 203, pp. 203-380, Jul 1964.