The influence on cognitive processing from the switches of shooting angles in videos of real-world events: An ERP study

Baolin Liu*, Zhongning Wang, Zhixing Jin, Guanjun Song

Department of Computer Science and Technology, Tsinghua University, Beijing 100084, PR China

**Article info**

**Abstract**

This work mainly focuses on the influence from switches of shooting angles in videos during the cognitive processing in the human brain. In the experiment we used the videos with switches of shooting angles as materials and compared the ERPs elicited by the switch frames and the non-switch frames in the videos, it was found that when subjects were asked to pay attention to the video contents, the switch frames would trigger P3a–RON waveforms, but no N400 waveform was found in the ERP results. This showed that when subjects were concerned with the video contents, the switches of shooting angles would distract their attention from the video contents, but as long as the semantic meaning of the videos were coherent, the switches of shooting angles would not lead to significant difficulties in semantic comprehension. At the same time, the experimental results also further proved that the P3a and RON generally reflect the processing of task-irrelevant visual stimuli.

© 2009 Elsevier Ireland Ltd. All rights reserved.

During the processing of the objective world in the human brain, 80% of the information is obtained through a visual channel. Video information plays an important role during the understanding of the world in the human brain. There are a lot of switches of shooting angles (switching the lens) in videos of real-world events, and these angle switches could influence the cognitive processing in the human brain, and therefore the research on cognitive processes of videos of real-world events would be of great significance to studying the cognitive mechanism in the human brain. Of several research methods available, the event-related potentials (ERPs) method is commonly used as it is a more mature and reliable experimental research method.

More static visual stimuli were used in the previous ERP studies on the cognitive processes in the human brain, but in recent years some studies using dynamic visual stimulation, such as video stimuli, were also reported [12,16,18,21,22]. At present, the researchers mainly focused on the cognitive processing of the inconsistent semantic meaning in video with a continuous movement. It has been proved that the latter semantic actions which were inconsistent with previous actions would evoke the N400 effect [18,22]. Many studies have proved that when there are words or sentences which are inconsistent with the context, or paragraph, an N400 effect would occur [6,8,9,11], so people generally considered that the N400 was related with the processing of semantic confusion in language comprehension. Later, some studies proved that when visual and auditory information were presented simultaneously, N400 effects would also be evoked [5]. Some studies on picture sequence indicated that when a picture which was inconsistent with the former and latter picture in content, there would also be an N400 effect [23]. Thus, it is considered that the N400 effect was not only related with the semantic integration in language comprehension, but also with that in processing of other types of stimuli.

In the 2003 studies on the N400 effect elicited by videos, Sitnikova et al. conducted an experiment on video materials [22]. They found that the video actions which were inconsistent with normal behavior would trigger a significant N400 effect, which was mainly distributed in the frontal region, and the duration of this N400 was longer than that evoked by pictures and languages (the video N400 sustained 600 ms). They considered that the timing of the appearance of target objects was somewhat more variable across the videos than static visual stimuli, and thus the N400 and N300 joined together. In 2007, Vincent and Reid conducted experiments on the video semantics [18]. They found that the video stimuli triggered a clear N400 waveform when the video clips contained some violations of people’s normal habits. These N400 effects generated from semantic inconsistencies were usually found in sequences of events. Combining these results, we believe that the human brain would integrate the external stimuli, and create some expectation for the following events. If the following events contradicted to the exception, the typical N400 effect would be evoked, representing the semantic integration processing.

© 2009 Elsevier Ireland Ltd. All rights reserved.
Another aspect which is worthy of our concern is the attention shift in the human brain. The studies on picture and sound sequence proved that the MMN, P3a, and RON (re-orienting negativity) waveforms may reflect the three stages of processing when the human brain encounters the interference in cognitive processing [1,4,7,19,20]. The researchers found that, in successive stimulation of sounds or pictures, task-irrelevant stimuli would evoke the MMN, P3a and RON components in sequence. The MMN reflects the subject’s observation of the task-irrelevant stimuli, the P3a reflects the processing of the task-irrelevant stimuli, and the RON reflects the human brain re-directing the cognitive resources to the task-relevant stimulus. We speculate that, when sequential pictures or videos are presented, the attention-related resources may also be affected in the human brain if a mismatched stimulation appeared in the sequence.

In addition, P200 component has also been discovered in the studies on video cognition. P200 is considered to be a typical waveform which reflects early attention in the human brain. It is usually related with phonological and orthographic processing [2,3,14]. In video studies, the latency of the P200 sometimes extends to 250 ms, expressed as P250, which is also the reflection of the early attention in cognitive processes. It is related with the complexity of video and the interaction between the visual and auditory information [12,17].

Usually there are switches of shooting angles in videos of real-world events, and they would affect the complexity and semantic information of the videos, thus the cognitive processing in the human brain would be influenced. In this study, we attempts to construct some simple silent continuous videos with switches of shooting angles, and corresponding picture sequences constructed by selecting the key frames which represented the shooting switches from the continuous videos mentioned above. By observing the P300, N400, RON and other ERP components, we will try to study the integration processing of the inconsistent semantic information elicited by switches of shooting angles in the human brain using the ERP method, moreover, to reveal the cognitive mechanism in the human brain.

We asked 21 students (11 males, 10 females, mean age 22 years (SD = 1.2)) from Tsinghua University to take part in our experiment. All participants were right-handed according to the Handedness Questionnaire (Chinese version) [15]. They were free of medication for at least one week before the experiment and had no history of neurological diseases. All of them had normal or corrected-to-normal vision and none of them were color blind. Written informed consent was obtained from each subject in a form approved by the Ethical Committee of the IHB and in agreement with the Declaration of Helsinki. Each participant was paid 35 yuan (RMB) per hour for his/her participation.

Video clips for experimental materials contained six different scenes, with consecutive actions and some switches of shooting angles. Each kind of scene was divided into two categories, namely: (1) original videos. These videos were not processed whatever and were defined as ‘Ori’ condition. (2) Picture sequence videos. The switch frames were selected from the original video, and a non-switch frame was selected between two switch frames. These frames would construct a picture sequence, and were presented in accordance with the order of the original video as a continuous video. These stimuli were defined as ‘Pic’ condition. In fact, this kind of stimulus was equivalent to a continuous video. Fig. 1 shows several switch frames and non-switch frames selected from the original video. Each picture was displayed for 960 ms, with an interval of 640 ms between two pictures. The lengths of the videos under the Ori condition were 8–21 s (mean length 14.6, SD = 4.32), and the lengths of the videos under the Pic condition were 12–27 s (mean length 18.0, SD = 5.90). The videos under the Pic condition contained 7–17 (mean amount 11.3, SD = 3.77) pictures.

In accordance with the requirements of our experiment, we complied with the following principles on video material selection to design a more standardized test: (1) we featured commonplace actions and sounds in the videos to ensure participants were familiar with the video topic and to prevent any interference with cognitive processing; (2) the videos do not contain emotional factors in order to avoid impact on the experimental results of subjects from emotional factors; (3) there were significant switches of shooting angles in the switch frame in comparison to the previous frame. There was no switch frame which was hard to distinguish, such as slow changing in shooting angles or a sudden zoom in and zoom out in shooting perspective; (4) the intervals between two switch frames lasted at least 1300 ms in the original videos to ensure the ERP of the former switch frame would not affect the following ERP; (5) the last switch frame appeared at least 1000 ms before the end of each video and the first switch frame appeared at least 1000 ms after the beginning of the video, to avoid any impact on the ERP waveforms from the appearance and the disappearance of the video clips. Each video clip would be presented 30 times, with a total of 360. For each subject, before the experiment, the 360 video clips would be pseudo-randomized, while videos from the same type (Ori or Pic) would not be played three times or more than three times consecutively.

All silent video clips were presented at 25 fps (frames per second), using the NTSC standard with a resolution of 720 × 480. They were edited with Adobe Premiere Pro. Frame-by-frame comparison approach was adopted to ensure that the frames selected were exactly the switches of shooting angle in the videos. By using the Windows Movie Maker software, the materials were compressed into the .wmv format with a smaller file size in order to avoid any time delays when the program read video files. Meanwhile, there were no significant drops in image quality. The video bit-rate was 1900 kbps with a 24 Hz sampling accuracy.

We conducted a pretest to score the emotional factor, switches of shooting angle and the familiarity of the videos, in order to ensure the experimental materials met our requirements. The degree of emotion was divided into 7 degrees [10,13], with 1 representing totally negative emotion, and 7 representing totally positive emotion. The degree of switches was divided into 7 degrees, with 1 representing no switches, and 7 representing significant switches of shooting angles. The familiarity of the videos was divided into 7 degrees, with 1 representing totally unfamiliarity and 7 representing very high familiarity. Another 20 volunteers from Tsinghua University were asked to take part in the pretest. The results were as follows: the average point of emotion score was 4.27 (SD = 0.6); the average point of switch was 6.58 (SD = 0.3); the average point of familiarity score was 5.83 (SD = 0.4).

Participants were seated in a comfortable chair in a soundproof and electrically shielded room. They were approximately 90–100 cm away from a computer screen, which was a 17 in. LCD monitor with a resolution of 1280 × 1024. A re-scaling was needed in order to fit the screen when the video clips were presented, and the aspect ratio displayed was 426 × 469 (2010) 370–374. A re-scaling was needed in order to fit the screen when the video clips were presented, and the aspect ratio displayed was 426 × 469. Video clips were presented in the center of the screen. The horizontal angle of view was about 6.74 ° and the vertical angle of view was about 4.5 °. Before the experiment, the 360 video clips were randomly divided into 6 groups with each group containing 60 clips. The presentation sequence of the 6 groups was randomized. There were proper rests between groups. Subjects were introduced to the content and the judgment tasks of the experiment, and there was also a stage of familiarity before the experiment. During the experiment, the background of the screen is black. Before each video was presented, a white cross would appear in the center of the screen, to prompt the subjects to focus their attention on the following video. The duration of the white cross was 800–1000 ms in order that subjects could not accurately predict the start time of each video. A judg-
The threshold of other artifacts were excluded manually from further analyses with baseline interval. Epochs contaminated by ocular movements or onset, with subtracting mean amplitude in a 200 ms pre-stimulus epoch from 200 ms before to 1000 ms after the critical stimulus posterior region. ERPs were calculated for each subject over an anterior region and CPz, CP3, CP4, Pz, P3, P4, P8, O1, Oz, O2. The horizontal EOG (HEOG) was recorded from two additional electrodes located outside of both eyes and the vertical EOG (VEOG) was monitored from two additional electrodes located above and below the left eye. Electrode AFz on the cap was selected as grounding. The bilateral mastoids were used as the recording reference, i.e., EEG data was referenced on-line to the linked mastoids. We used Electrode gel in order to get an impedance of less than 5 kΩ. The signals were digitized at 1000 Hz and amplified with a band pass from 0.01 to 70 Hz.

Mean amplitudes of N170, P3a and RON were obtained over the time windows of 150–200, 230–280 and 400–500 ms respectively. Repeated measure ANOVAs were performed with three factors: Condition (switch frames in the Pic videos vs. non-switch frames in the Pic videos), Region (anterior vs. posterior) and Electrode (6 levels), with Fz, F3, F4, FCz, FC3, FC4, FT8, T7, C3, Cz, C4, T8, TP7, CP3, CPz, CP4, TP8, P7, P3, Pz, P4, P8, O1, Oz, O2. The horizontal EOG (HEOG) was recorded from two additional electrodes located outside of both eyes and the vertical EOG (VEOG) was monitored from two additional electrodes located above and below the left eye. Electrode AFz on the cap was selected as grounding. The bilateral mastoids were used as the recording reference, i.e., EEG data was referenced on-line to the linked mastoids. We used Electrode gel in order to get an impedance of less than 5 kΩ. The signals were digitized at 1000 Hz and amplified with a band pass from 0.01 to 70 Hz.

Mean amplitudes of N170, P3a and RON were obtained over the time windows of 150–200, 230–280 and 400–500 ms respectively. Repeated measure ANOVAs were performed with three factors: Condition (switch frames in the Pic videos vs. non-switch frames in the Pic videos), Region (anterior vs. posterior) and Electrode (6 levels), with Fz, F3, F4, FCz, FC3, FC4 representing the anterior region and CPz, CP3, CP4, Pz, P3, P4 representing the posterior region. ERPs were calculated for each subject over an epoch from 200 ms before to 1000 ms after the critical stimulus onset, with subtracting mean amplitude in a 200 ms pre-stimulus baseline interval. Epochs contaminated by ocular movements or other artifacts were excluded manually from further analyses with the threshold of ±100 μV. Two subjects’ data were excluded from further analysis because of their low artificial free data (lower than 65%). The experimental data with incorrect judgments were excluded too. Finally, the average proportions of the trials entered into the final analysis for different conditions were respectively as follows: 86.3% for the switch frames under the Ori condition, 87.7% for the non-switch frames under the Ori condition, 84.3% for the switch frames under the Pic condition and 85.4% for the non-switch frames under the Pic condition. The Greenhouse–Geisser correction was applied to all analyses when evaluating statistical effects with more than one degree of freedom in the numerator.

In our experiment, subjects were asked to judge the questions presented after the videos. The behavioral data showed that the judgment accuracy of videos was 85.2% (SD = 3.8%) under the Ori condition, and 81.3% (SD = 4.5%) under the Pic condition. Overall, the judgment accuracy of the two kinds of videos was high, which basically indicated that subjects paid attention to the video contents. The judgment accuracy of videos under the Pic condition was slightly lower than that of videos under the Ori condition [F(1,18) = 3.352, p < .01, MS = 78.293], which indicated that a continuous presentation of picture sequence could cause certain difficulties when the human brain determined the content of this kind of video, due to the ever-changing pictures.

As shown in Fig. 2, under the Pic condition the switch frames and the non-switch frames could elicit a negative wave in the frontal and central regions at about 170 ms. There was no significant difference between these two waves [F(1,18) = 3.652, p = .71, MS = 216.810] nor interaction between Condition and Region [F(1,18) = 1.834, p = .75, MS = 130.329]. In the parietal region the switch frames and the non-switch frames both elicited significant P3a and P3b components, with the time windows of 230–280 and 400–500 ms respectively. Under the Pic condition, the switch frames evoked a larger P3a wave [F(1,18) = 7.099, p = .015, MS = 451.333] in the time window of 230–280 ms, and a smaller P3b in the time window of 400–500 ms. As a result a negative wave was elicited compared with the ERPs aroused by non-switch frames [F(1,18) = 3.109, p = .094, MS = 35.842]. After 600 ms, the trends of ERP waves aroused by the switch frames and the non-switch frames were basically the same, although there were slight differences in amplitude [F(1,18) = 3.912, p < .05, MS = 51.392]. The trends of the waveforms by the switch frames under the Ori condition and by the switch frames under the Pic condition are basically the same, while the ERP amplitudes elicited by the switch frames under the Pic condition are larger [F(1,18) = 4.391, p < .05, MS = 40.934].
Fig. 2. The grand average ERPs in response to the switch frames and non-switch frames under the Ori/Pic condition were shown in left part of the figure. Under the Pic condition, the switch frames elicited a larger P3a in the time window of 230–280 ms, and a RON effect in the time window of 400–500 ms compared with the non-switch frames. There is no significant ERP elicited by the non-switch frames under the Ori condition. The ERPs aroused by the switch frames under the Pic condition and under the Ori condition are similar between each other, and the trends are approximately the same, but the switch frames under the Pic condition elicited a larger ERP waveform compared with that under the Ori condition. The topographic maps of the ERP differences between the switch frames under the Pic condition and the non-switch frames under the Pic condition were plotted in the right part of the figure.

In our results, the trends of ERP waveforms elicited by the switch frames under the Ori condition and by the switch frames under the Pic condition are basically the same, with a little difference in amplitude. This result indicated that the switch frames extracted from Ori videos under the Pic condition could basically restore the ERP characteristics of the switch frames under the Ori condition. So we could approximate cognitive processing of switch frames and non-switch frames under the Ori condition by comparing the ERP waves elicited by the switch frames and the non-switch frames under the Pic condition.

In 2003, Sitnikova and Kuperberg mentioned that the usage of video switches might measure the latency of semantic related N400 more accurately, thereby eliminating the aliasing of N300 and N400. The video materials used by Sitnikova and Kuperberg contained semantic confusion between former and latter parts of the videos [22]. In our experiment, we found that the switch frames could elicit a negative wave compared with the non-switch frames in the time window of 400–500 ms. However, the video materials selected in our experiment were coherent in semantics, there was no semantic confusion within them. Thus we speculated that this negativity might not be the N400 or N300 waves reflecting the processing of semantic confusion.

The switches of shooting angles in continuous videos could influence the cognitive processing. The switches could distract the attention from what we focused on, and lead to the change of ERP waves. Some earlier studies have found that some task-irrelevant stimuli could distract the attention in the human brain and elicit the MMN (mismatch-negativity), P3a and RON (re-orienting negativity) waves. These three waves usually stand for the procedure of noticing the task-irrelevant stimuli, processing the stimuli and re-orienting the attention to the factors which should be focused on [1,20]. These waves are considered to be related with an unconscious attention switch, and regarded as the typical waveforms of task-irrelevant stimuli processing. In our experiment, subjects were asked to pay attention to the contents of the videos and make judgments on the questions concerning the video contents. When there were switches of shooting angles, these switches might be a kind of task-irrelevant stimuli, and elicited the serial MMN-P3a-RON waves.

In general, MMN waveform tends to occur in the auditory attention switching experiments, and thus MMN component is generally considered to be an automatically triggered, unconscious processing of auditory change detection [1,19,20]. In our experiment, the switch frames and the non-switch frames could elicit N170s, but
there were no significant differences between them. So no obvious MMN effect was found. This result was consistent with what Berti and Schröger found in their visual attention shifting study in 2001 [1]. P3a can also be regarded as being related with attention resources to deal with the task-irrelevant stimulation from the switches. In the time window of 400–500 ms, the switch frames triggered a negative wave compared with the non-switch frames under the Pic condition. We believed that this wave was the RON wave which reflects the process of re-directing attention. The latencies of RON waves found in earlier studies were usually 400–600 ms [20], while the latency of RON in our experiment was relatively shorter. This might be because the degree of distracting attention by switches of shooting angles is relatively smaller compared with other kinds of task-irrelevant stimuli, so the human brain could process these stimuli faster and re-orient its attention to task-related stimulation more quickly.

In this paper we studied the influence from the switches of shooting angles in videos on cognitive processing. By using semantic continuous videos with switches of shooting angles, we compared the ERP waves elicited by the switch frames and the non-switch frames, and found that when subjects were asked to focus on the video contents, the switches of shooting angles in the videos would distract the attention and elicit the P3a-RON waves, but no N400 effect was found. The P3a-RON waves reflect the procedure of processing the task-irrelevant stimuli and re-orienting the attention resources to what should be focused on. These results indicated that the switches of shooting angles in the videos have less of an influence on semantic meanings of the videos, but the attention resources would be affected in the human brain. The experimental results further verified that the P3a and RON might be common ERP waves which reflect the attention shift by visual stimuli.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (No. 90820304). The authors would like to thank the Laboratory of Neural Engineering (Tsinghua University) for providing Neuroscan Synamps system.

References