Neuroscience 194 (2011) 19–26

THE INFLUENCE OF MATCHING DEGREES OF SYNCHRONOUS AUDITORY AND VISUAL INFORMATION IN VIDEOS OF REAL-WORLD EVENTS ON COGNITIVE INTEGRATION: AN EVENT-RELATED POTENTIAL STUDY

B. LIU,* Z. WANG AND J. LI

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

Abstract—In this article, we aim to study the influence of matching degrees of synchronous natural auditory and visual information on cognitive integration. Videos with matched, moderately matched, and mismatched audio-visual information were used as stimuli. The results showed that videos with moderately matched audio-visual information could elicit N400, P600, and late negativity (LN) effects, while videos with mismatched audio-visual information could elicit N400 and late negativity effects as compared with those with matched audio-visual information. It was further proven that N400 might reflect the connection process during multisensory integration, and P600 was more related to the evaluation process on the matching degrees of the audio-visual information in videos. Late negativity under the mismatched condition might be the combination of late frontal negativity (LFN) and late posterior negativity (LPN), which reflected the attention reallocating process and the recognition process, while late negativity under the moderately matched condition might be the LPN, which was related to the recognition process in the human brain. It was demonstrated that cognitive integration of synchronous audio-visual information would be modulated by different matching degrees of audio-visual information as indexed by different event-related potential (ERP) effects.

Key words: video, multisensory integration, matching degree, ERP.

In real life, 80% of natural information is received by the human brain from auditory and visual senses. Through integration, the brain would get a comprehensive recognition of the real world. During this process, semantics plays an important role, and multisensory integration would be modulated by the semantically matching degrees between auditory and visual information. Therefore, study on the influence of matching degrees of synchronous audio-visual information on cognitive integration is of great importance.

In unisensory studies, Sitnikova et al. (2003) used natural videos with congruent and incongruent actions to study the semantic processing in the human brain. They found that incongruent actions would elicit N400-like negative effect when the latter actions were incongruent with the former ones. This N400-like effect was quite similar to the N400 effect found in language studies (Kutas and Hillyard, 1980; Hagoort et al., 2004; Kuperberg et al., 2003b; Liu et al., 2009a, 2010b,c, 2011a,c), which suggested that the semantic processes of natural information and language might be similar. Reid and Striano (2008) found that latter mismatched actions in continuous action sequence would also elicit N400 effect compared with the former ones. This result provided further evidence that the N400 effect could reflect semantic processing of natural information in the human brain.

Later, Sitnikova et al. (2008) conducted further research to study semantic processing of videos. They found that totally semantically unexpected endings of videos would elicit N400 effect, while goal-related unexpected endings of videos would elicit both N400 effect and a late positive wave, which were similar to the N400-P600 effects found in language studies (Liu et al., 2009a, 2010b,c, 2011a,c). They considered that there were two independent semantic processing mechanisms during natural information cognition. The first mechanism, reflected by the anterior N400-like negativity, maps the incoming information onto the connections of various strengths between concepts in semantic memory. The second mechanism, reflected by the posterior late positivity, evaluates the incoming information against the discrete requirements of real-world actions (Sitnikova et al., 2008).

Van Petten and Rheinfelder (1995) studied the cognitive processes of natural sounds and speeches. They found that semantically mismatched natural sounds and speech would both elicit similar N400 effect.

In recent years, synchronous audio-visual integration studies have also been reported (Liu et al., 2009b; Puce et al., 2007; Molholm et al., 2004). Molholm et al. (2004) found that when pictures and sounds were presented synchronously, mismatched auditory and visual elements could elicit a more negative going response than those by matched auditory and visual elements. They considered this effect to be related with the semantic congruency between the simultaneously presented auditory and visual elements, which could be reflected by N400 effect. This result thereby indicates that multisensory semantic integration is similar to the unisensory semantic processing.

However, Puce et al. (2007) used human face, monkey face, and house as visual stimuli, and human voices, mon-

*Corresponding author. Tel: +86-10-62781789; fax: +86-10-62771138.
E-mail address: liubaolin@tsinghua.edu.cn (B. Liu).

Abbreviation: ERP, event-related potential; LFN, late frontal negativity; LN, late negativity; LPN, late posterior negativity; VA*, moderately matched condition; VA−, mismatched condition; VA+, matched condition.

0306-4522/11 $ - see front matter
doi:10.1016/j.neuroscience.2011.08.009

key voices, and house sounds as auditory stimuli to study the cognitive processing of matched and mismatched pictures and sounds. Significant P400 effect was found when the semantic information of pictures and sounds were inconsistent (Puce et al., 2007). No N400-P600 waveforms were found in their experiment. Therefore, synchronous multisensory semantic integration is still worth further study.

In the studies on multisensory semantic integration mentioned previously, only semantically matched and mismatched audio-visual stimuli were used as experimental materials. In fact, different matching degrees of audio-visual information might also affect the multisensory integration processes. Therefore, we would construct continuous videos with different semantically matching degrees of audio-visual information (matched, mismatched, or moderately matched) as experimental materials in our experiment. Thereafter, we would observe some possible ERP effects, such as N400 and P600, to study the influence of matching degrees of synchronous natural audio-visual information on multisensory integration. Furthermore, we would try to conclude the cognitive mechanisms of multisensory integration processing in the human brain.

EXPERIMENTAL PROCEDURES

Participants

Twenty students from Tsinghua University (10 females and 10 males, mean age 23.2 years [SD=2.2]) took part in the experiment. They had normal hearing and normal or corrected-to-normal vision. None of them were color blind. All participants had no history of neurological diseases, and were free of medication for at least 1 week before the experiment. All were told that our experiment was conducted under the Declaration of Helsinki, and all signed the Researchers' Consent Form. They were judged to be right-handed according to the Edinburgh Handedness Inventory (Oldfield, 1971). Each participant was paid 35 Yuan (RMB) per h for his/her participation.

Materials

In this experiment, six different kinds of original videos were selected. Each video represented a simple natural scene, and contained a critical action. The six scenes were as follows: (1) a wine glass falling from a height and shattering; (2) a person with his back towards the screen going out and closing the door; (3) a cue-stick striking a white billiard ball; (4) a large display of festive fireworks exploding in the sky; (5) a water droplet landing with a splash; and (6) a small ball on a horizontal board colliding with another ball.

For each video, we prepared three kinds of natural sounds that were temporally synchronous with the critical actions in the videos: (1) a sound that is totally semantically matched with the critical action in the video; (2) a sound that is totally mismatched with the critical action in the video; (3) a sound that is moderately matched with the critical action in the video. Take the video of the water droplet for example. The matched condition, VA+, is defined if a matched sound is presented at the moment when the water droplet lands into the body of water. The mismatched condition, VA−, is defined if a mismatched sound, i.e., a sound of knocking gong, appears. The moderately matched condition, VA*, is defined if an action-related but not the matched sound, i.e., a sound like that of a piece of cobblestone falling into water, is presented (see Fig. 1).

We complied with the following principles in the construction of the experimental materials to conduct a more normative experiment: (1) commonplace actions and sounds in the videos were featured, and emotional elements were also avoided in order to prevent familiarity degree and emotional factor from influencing cognitive processing; (2) the videos were clear and continuous, and there was no switch of shooting angles (Liu et al., 2010d); (3) the critical actions were instantaneous actions, as shown in Fig. 1, with the critical action being the moment when the water droplet lands into the body of water; (4) participants would be guided by the context of the video to have an expectation consistent to the following critical action in the moment preceding the critical action; (5) natural sound was generated by critical action, and there is no other sound before or after that; (6) it is easy to distinguish the matching degrees between the semantic meanings of the critical action and the sound; (7) the onset of the critical action was at least 1000 ms after the beginning of the video and 1000 ms before the end of the video, in order to avoid influence from the beginning and end of the video clip. Each video clip was presented 20 times, amounting to 360 (6×3×20) video clips in total.

The original video clips were edited in Adobe Premiere Pro (Adobe Systems Inc., San Jose, CA, USA) with NTSC standard, and the resolution used was 720×480. They were presented with 25 fps (frame per second). All sounds were properly edited in Cool Edit Pro 2.1 (Adobe Systems Inc.) with normalized volume. After editing, the sounds were re-integrated with the video clips by Adobe Premiere Pro, during which we adopted a frame-by-frame comparison approach to ensure that the critical actions and the sounds were synchronized. For example, in the water droplet scene, we searched the video for the exact frame in which the droplet landed into the water body, and then added the matched, mismatched, or moderately matched sound. Thereafter, all the

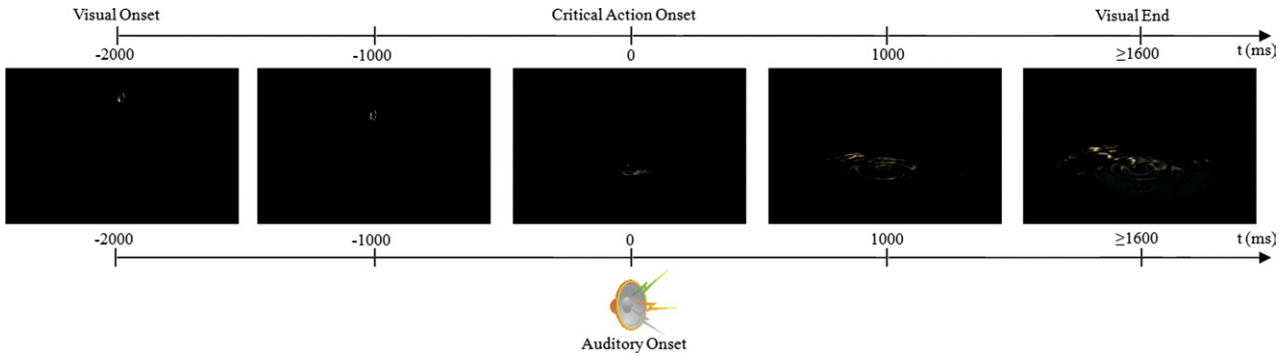

Fig. 1. Illustrational video frames of the experimental stimuli. The critical action onset and auditory onset are synchronous. For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.