Biological Psychology 87 (2011) 200–208

Contents lists available at ScienceDirect

Biological Psychology

journal homepage: www.elsevier.com/locate/biopsycho

Neural mechanisms for the effect of prior knowledge on audiovisual integration

Qiang Liu[a,b], Ye Zhang[a,b], Jennifer L. Campos[c,d], Qinglin Zhang[a,b,*], Hong-Jin Sun[c,**]

[a] *Key Laboratory of Cognition and Personality, Ministry of Education, Southwest University, Chongqing 400715, China*
[b] *School of Psychology, Southwest University, Chongqing 400715, China*
[c] *Department of Psychology, Neuroscience and Behaviour, McMaster University, Hamilton, Ontario L8S 4K1, Canada*
[d] *iDAPT, Department of Technology, Research and Development, Toronto Rehabilitation Institute, Toronto, Ontario M5T 3M4, Canada*

ARTICLE INFO

Article history:
Received 21 May 2010
Accepted 15 February 2011
Available online 4 March 2011

Keywords:
Prior knowledge
Multisensory integration
Event-related potentials
Audiovisual integration

ABSTRACT

Converging evidence indicates that prior knowledge plays an important role in multisensory integration. However, the neural mechanisms underlying the processes with which prior knowledge is integrated with current sensory information remains unknown. In this study, we measured event-related potentials (ERPs) while manipulating prior knowledge using a novel visual letter recognition task in which auditory information was always presented simultaneously. The color of the letters was assigned to a particular probability of being associated with audiovisual congruency (e.g., green = high probability (HP) and blue = low probability (LP)). Results demonstrate that this prior began affecting reaction times to the congruent audiovisual stimuli at about the 900th trial. Consequently, the ERP data was analyzed in two phases: the "early phase" (<trial 600) and the "late phase" (>trial 900). The effects of prior knowledge were revealed through difference waveforms generated by subtracting the ERPs for the congruent audiovisual stimuli in the LP condition from those in the HP condition. A frontal-central probability effect (90–120 ms) was observed in the early phase. A right parietal-occipital probability effect (40–96 ms) and a frontal-central probability effect (170–200 ms) were observed in the late phase. The results suggest that during the initial acquisition of the knowledge about the probability of congruency, the brain assigned more attention to audiovisual stimuli for the LP condition. Following the acquisition of this prior knowledge, it was then used during early stages of visual processing and modulated the activity of multisensory cortical areas.

1. Introduction

To perceive a complex external environment, our brains make use of multiple cues derived from different sensory modalities (e.g., vision, touch and audition). Typically, cues from different sensory systems are efficiently merged to form a unified and robust percept. This process is referred to as multisensory integration (see Ernst and Bülthoff, 2004 for a review). Converging evidence from human behavioral research has demonstrated that stimuli from two or more sensory modalities presented in close spatial and/or temporal proximity can have a facilitative effect on behavioral performance. Specifically, multimodal stimulation leads to faster detection times and more accurate discrimination performance compared to the constituent unimodal stimuli (Frassinetti et al., 2002; Frens et al., 1995; Hershenson, 1962; Loveless et al., 1970). Recent modeling and behavioral research has demonstrated that the brain integrates cues from multiple sensory modalities by taking a linear combination of the individual sensory estimates with each sensory estimate weighted according to its relative reliability (Alais and Burr, 2004; Ernst and Banks, 2002; Gepshtein and Banks, 2003; Helbig and Ernst, 2007; Bresciani et al., 2006; Jacobs, 1999; Knill and Saunders, 2003; van Beers et al., 1999). The reliability of a cue is inversely proportional to the variance of the noise distribution associated with the unimodal sensory estimate. This integration therefore conforms to a statistically optimal rule.

In addition to relative reliability, previous behavioral studies have shown that prior knowledge likely also affects the weights of the cues from different modalities. For example, previous studies have investigated whether tactile feedback affects the integration of visual texture and visual disparity when estimating slant (Ernst et al., 2000) and whether tactile feedback affects the integration of visual texture or visual motion when estimating depth (Atkins et al., 2001). Specifically, in these experiments, during the training phase, tactile feedback was always consistent with one of the visual cues, while the other visual cue was randomly varied. After training, the results showed that the visual cue that was reinforced by tactile feedback played a much more dominant role when estimat-

* Corresponding author at: School of Psychology, Southwest University, Chongqing 400715, China. Tel.: +86 23 6825 3939; fax: +86 23 6825 3304.
** Corresponding author. Tel.: +1 905 525 9140x24367.
E-mail addresses: lq780614@163.com (Q. Zhang), sunhong@mcmaster.ca (H.-J. Sun).

0301-0511/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.biopsycho.2011.02.006

ing slant. Accordingly, these findings suggest that when the sensory estimates from one modality are not always consistent with others, the brain may consider them to be unstable. To increase the weighting of the more stable estimates, prior knowledge about the uncertainty of sensory information could be stored and used (Ernst and Bülthoff, 2004). Moreover, Atkins et al. (2001) demonstrated that participants could adapt their cue integration strategies in a context-dependent manner based on correlations between haptic and visual cues during the training phase. In their experiment, visual motion and haptic cues were always congruent when the object was red, but visual texture and haptic cues were always congruent when the object was blue. After having learned this relationship over repeated trials, the weighting of specific visual cue (motion or texture) would change depending on the object color. This finding indicates that the brain is able to pick up multiple types of prior knowledge simultaneously, and extract the most appropriate feature based on the context.

Although there have been several recent behavioral studies evaluating the specific effects of prior knowledge on cue-integration, the neural mechanisms underlying these processes remain poorly understood. For example, the stage of cognitive processing during which prior knowledge impacts multisensory integration is currently not known. Using only behavioral data, it is not possible to assess the temporal properties of different stages of cognitive processing. Event-related potential (ERP) recording techniques, however, provide a means by which to evaluate the timing of perceptual and cognitive processes prior to behavioral responses. Using this technique, the electrical activity of the brain is time-locked to the presentation of an external stimulus. Thus, ERP data allows for more precise statements to be made about the time course of activation during different stages of multisensory processing.

The purpose of the present study was, therefore, to investigate the spatiotemporal patterns of brain activation during the different phases of multisensory integration using high-density (64-channel) ERP recordings. Previous multisensory ERP studies have extracted the effects of multisensory integration through comparing ERPs elicited by combined multisensory stimulus with summed unisensory ERPs. Using this approach, previous research has found that multisensory interactions may occur at sensory-specific areas (starting at approximately 30–50 ms post-stimulus) and at multiple scalp areas (during 100–250 ms post-stimulus) (Molholm et al., 2002; Fort et al., 2002; Giard and Peronnet, 1999; Klucharev et al., 2003; Foxe et al., 2000; Murray et al., 2005; Santangelo et al., 2008). Moreover, several studies have reported that audiovisual integration is modulated by attention to audiovisual stimuli (Talsma and Woldorff, 2005) and is differentially modulated by attention to one particular modality or both (Talsma et al., 2007; Latinus et al., 2010). Multisensory fMRI studies have revealed that areas in which multisensory interactions occur include, the superior temporal sulcus (STS), intraparietal sulcus (IPS), frontal cortex, superior colliculus, basal ganglia, and putamen (for reviews, see Macaluso and Driver, 2005; Macaluso, 2006; Calvert and Thesen, 2004; Ghazanfar and Schroeder, 2006; Thesen et al., 2004). However, the experimental paradigms adopted by previous ERP and fMRI studies have not directly manipulated specific aspects of prior knowledge.

To manipulate prior knowledge in the current study, we designed a new experimental paradigm using visual and auditory letter stimuli. The experimental task required participants to identify a letter based only on visual information. Auditory and visual letters were always presented simultaneously, but the color of the visual letters was either green on some trials or blue on other trials. During the practice phase of the experiment, the visual and auditory information were always congruent. However, in the experimental phase we manipulated the probability with which the visual and auditory information were congruent, which was contingent on the color of the stimulus. For example, for some participants, green letters were associated with a high proportion (HP) of congruent cue trials (100%) and blue letters were associated with a low proportion (LP) of congruent cue trials (30%).

In light of previous research findings that have shown a stable audiovisual integration effect for congruent audiovisual letter identification (Raij et al., 2000; van Atteveldt et al., 2004, 2007; Liu et al., 2007; Blau et al., 2008), we predicted that participants would optimally integrate visual and auditory letter cues when they were congruent. Furthermore, according to the findings of Atkins et al. (2001), we predicted that, for the HP and LP conditions, knowledge about the relation between the letter color and the probability of congruent audiovisual letters would be acquired gradually and used in a context-dependent manner. This would then result in a modulation of audiovisual integration, which would be demonstrated by the behavioral and ERP data for the congruent audiovisual stimuli as a function of the frequency of congruent audiovisual letters. Further, spatiotemporal analyses of these ERP components will provide further insights into the electrophysiological mechanisms underlying the effects of prior knowledge on multisensory integration.

2. Materials and methods

2.1. Participants

A total of 14 undergraduates (seven females and seven males with a mean age of 22.3 years old) participated in the experiment as paid participants. All participants were right-handed, native Chinese speakers with normal or corrected-to-normal vision. All participants gave their informed written consent before participating in the study. This research was approved by the Research Ethics Committee of Southwest University of China and was conducted in accordance with the Declaration of Helsinki.

2.2. Stimuli

The stimuli consisted of graphemic representations of capital letters and speech sounds of two letters, B and E. The visual stimulus was displayed in blue or green against a gray background in the center of a 17-inch CRT monitor screen (positioned 70 cm away from the viewers' eyes) with a visual angle of 0.7 degrees using the font type "Song", which is quite similar to "Times new roman". The display had a screen resolution of 1024×768 and a screen refresh rate of 85 Hz. Speech sounds were digitally recorded (sampling rate 44.1 kHz, 16 bit quantization) and were produced by a female speaker. The speech sounds of all letters were identified correctly 100% of the time by all participants. The visual stimulus was 60 ms in duration. The speech sound had the same stimulus duration (including 8 ms of rise/fall times) and was delivered through Sennheiser HDR-110 headphones. The graphemic representations and speech sounds were presented synchronously and consisted of either the same content (congruent) or different content (incongruent). The congruent audiovisual stimuli included visual /B/ + auditory /B/ and visual /E/ + auditory /E/. The incongruent audiovisual stimuli include visual /B/ + auditory /E/ and visual /E/ + auditory /B/.

2.3. Procedure

Participants were seated in a dark, sound-attenuating room and were given instructions describing the task. Each trial was performed in the following sequence. A central fixation point was first presented for 500 ms followed by a blank gray screen for 300–500 ms. The bimodal stimulus was then presented for 60 ms after which a blank gray screen was presented until the participant pressed a key. The participants were instructed to press the button "1" if they detected a visual presentation of the letter B of any color, or press the button "2" if they detected a visual presentation of the letter E of any color. They were instructed to only report the visual letters irrespective of the simultaneously presented auditory letters. Reaction time and electrical brain activity were recorded. In the formal experiment, the proportion of congruent and incongruent visual-auditory stimuli was either a high (HP, with a probability of congruent audiovisual stimuli of 100%) or low (LP, with a probability of congruent audiovisual stimuli of 30%). With the equal probability of target color being blue or green, for half of the participants, the blue visual stimuli were completely congruent with the auditory stimuli (HP), while only 30% of the green visual stimuli were congruent with the auditory stimuli (LP). For the other half of the participants, the green visual stimuli were completely congruent with the auditory stimuli (HP), while only 30% of the blue visual stimuli were congruent with the auditory stimuli (LP). Each participant completed total of 1500 trials,