Neural mechanisms for the effect of prior knowledge on audiovisual integration

Qiang Liu a,⁎, Ye Zhang a,⁎, Jennifer L. Campos c,⁎⁎, Qinglin Zhang a,⁎, Hong-Jin Sun c,⁎⁎

⁎ Key Laboratory of Cognition and Personality, Ministry of Education, Southwest University, Chongqing 4000715, China
⁎⁎ School of Psychology, Southwest University, Chongqing 4000715, China
© Department of Psychology, Neuroscience and Behaviour, McMaster University, Hamilton, Ontario L8S 4K1, Canada
⁎⁎ iDAPT, Department of Technology, Research and Development, Toronto Rehabilitation Institute, Toronto, Ontario M5T 3M4, Canada

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 Bülböf, 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., 2000) and whether tactile feedback affects the integration of visual texture and visual disparity when estimating slant (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-

© 2011 Elsevier B.V. All rights reserved.
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 Bülböth, 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 stimuli 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 Woldorf, 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. 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 South-west 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 HDP-800 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 /E/ and visual /E/ + auditory /B/. 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,
with 750 trials for each probability condition (each color). The participants were not told about the different probabilities of experiencing congruent audiovisual stimuli before the procedure.

A practice block was performed prior to the formal experiment. In the practice block, the visual stimuli of two colors were completely congruent with the auditory stimuli. Each participant completed 40 trials, in which there were 20 trials for each color letter condition. Moreover, at the end of the formal experiment, the participants were asked to judge which color of printed letter was related to a higher probability of congruent audiovisual stimuli.

2.4. Analysis of behavioral data

Considering that acquiring prior knowledge is a gradual process, it would be predicted that the observed effects due to the “probability of congruency” prior would change across the time course of the experiment. Therefore, in order to reveal characteristics of this gradual learning process, we divided the 1500 trials into 5 blocks, each with an equal number of trials (blocks 1, 2, 3, 4, and 5 according to the time sequence). We then analyzed the difference between the average response times (RTs) to the congruent audiovisual stimuli under the two probability conditions in each block. Finally, a paired t-test was used to compare the difference between the RTs for these two probability conditions.

2.5. Electrophysiological (EEG) recording and analysis

Electrical brain activity was recorded from 64 scalp sites using tin electrodes mounted on an elastic cap (Brain Products), with references on the left mastoid. Vertical electrooculogram (EOG) recordings were obtained using electrodes placed above and below the left eye. The total inter-electrode impedance was maintained below 5 kΩ. The EEG and EOG were amplified and bandpass filtered with an analog filter of 0.1–100 Hz, at a sampling rate of 500 Hz. All signals were re-referenced off-line to an average of the mastoids and were bandpass filtered (0.1–30 Hz). The EEG and EOG were epoched off-line into 1000 ms periods including a 100 ms pre-stimulus baseline. Trials with eyeblinks (vertical EOG amplitudes exceeding ±100 μV), horizontal eye movements (horizontal EOG amplitudes exceeding ±25 μV), or other artifacts (a voltage exceeding ±100 μV at any electrode location relative to baseline) were excluded from the analysis.

In order to effectively capture the ERP components related specifically to prior knowledge, we divided the experiment into two different phases based on the behavioral results. Since a significant difference between the RTs for the HP and LP conditions was only observed during blocks 4 and 5 (i.e. the last 600 trials), we designated the first 600 trials as the “early phase” and the last 600 trials as the “late phase”. We expected that the prior knowledge associated with the two probability conditions would be evidenced by differences in the ERP waveforms for the congruent audiovisual stimuli. For each phase, ERPs for the congruent audiovisual stimuli (AVc) in the HP and LP probability conditions were individually averaged.

The effects of the probability condition on the processing of congruent audiovisual stimuli were first analyzed by comparing the ERP components involved in the difference waveforms (AVcHP − AVcLP) in the two phases. The difference waveforms (D = AVcHP − AVcLP) were obtained by subtracting the responses for the LP conditions from the responses for HP conditions. Significant effects (amplitude of the difference waveforms compared to zero) were assessed using a Student’s t-test (2-tailed), computed for each time window at each electrode in the two phases.

On the basis of the t-tests and the ERP topographical maps, the following 11 posterior sites were selected for statistical analysis of ERP data during the 40–96 ms interval [CP1, CP2, CPz, P1, P2, P3, P4, PO3, POz, PO4]: 9 electrode sites were selected for the statistical analysis of ERP data during the 90–120 ms interval [F1, Fz, F2, FC1, FC2, FCz, C1, C2, Cz]; 10 electrode sites were selected for the statistical analysis of ERP data during the 170–200 ms interval [F1, Fz, FC1, FC2, FCz, C2, AF3, AF4]. Mean amplitudes of ERP waveforms were analyzed using a three-way, repeated-measures analysis of variance (ANOVA). The factors included, electrode site, phase (early phase and late phase) and probability condition (HP and LP). For all ANOVAs, p-values were corrected for deviations according to Greenhouse Geisser when the degrees of freedom were more than one.

3. Results

3.1. Behavioral performance

Before the formal analysis, trials were rejected if RTs were less than 100 ms or more than two standard deviations away from the participant’s mean RT. Across the sequence of five blocks, the mean accuracies and standard deviations for the congruent audiovisual stimuli in the HP and LP conditions and for incongruent audiovisual stimuli in the LP conditions are shown in Table 1. The results of the paired t-test analysis show that the difference between accuracies for the HP and LP conditions was not significant for all blocks. The difference between accuracies for the congruent and incongruent audiovisual stimuli in LP conditions was significant only for block 1 \([t(1, 13) = 2.584, p < .05]\). The difference between accuracies for the congruent audiovisual stimuli in HP conditions and incongruent audiovisual stimuli in LP conditions was significant only for block 1 \([t(1, 13) = 2.765, p < .05]\) and block 2 \([t(1, 13) = 2.908, p < .05]\).

Across the sequence of five blocks, the mean RTs and standard deviations for the congruent audiovisual stimuli in the HP and LP conditions and incongruent audiovisual stimuli in LP conditions are shown in Table 2. The results of the paired t-test analysis indicate that the difference between RTs for the HP and LP conditions for the congruent audiovisual stimuli was significant for block 4 \([t(1, 13) = -3.162, p < .01]\) and block 5 \([t(1, 13) = 12.04, p < .01]\). The difference between RTs for the congruent and incongruent audiovisual stimuli in the LP conditions was significant for all blocks \(t(1, 13) = 8.19, p < .001\), block 2: \([t(1) = 7.85, p < .001]\), block 3: \([t(1) = -2.80, p < .05]\), block 4: \([t(1) = 2.47, p < .05]\), block 5: \([t(1) = 2.79, p < .05]\). The difference between RTs for the congruent audiovisual stimuli in the HP conditions and incongruent audiovisual stimuli in the LP conditions was significant for all blocks \(t(1, 13) = 9.24, p < .001\), block 2: \([t(1) = 7.74, p < .001]\), block 3: \([t(1) = 4.01, p < .01]\), block 4: \([t(1) = 4.66, p < .001]\), block 5: \([t(1) = 4.72, p < .01]\).

After having completed the experiment, all 14 participants indicated that they were unaware of a higher probability of congruent audiovisual letters associated with a certain letter color.

3.2. ERP waveforms analyses

The ERP waveforms evoked by the congruent audiovisual stimuli in the HP and LP conditions during the two phases (early phase and late phase) are shown in Fig. 1. A Student’s t-test comparing the amplitudes of the difference waveforms AVcHP − AVcLP against zero at each latency (see Fig. 2), showed that an effect of the probability manipulation occurred in three periods: (1) 90–120 ms in the early phase, (2) 40–96 ms in the late phase, and (3) 170–200 ms in the late phase.

3.2.1. Early phase: probability effects during the 90–120 ms interval

In the latency window of 90–120 ms, the results of the ANOVA showed significant main effects for probability \([F(1, 13) = 14.493, p < .01]\) and probability condition \([F(1, 13) = 14.812, p < .01]\), as well as a significant phase × probability condition interaction effect \([F(1, 13) = 7.246, p < .05]\). A simple effects analysis found a significant main effect of probability \([F(1, 13) = 30.52, p < .001]\) in the early phase, but not in the late phase \([F(1, 13) = .12, p > .05]\). A significant main effect was observed for phase in both the HP condition \([F(1, 13) = 9.27, p < .01]\) and the LP condition \([F(1, 13) = 14.96, p < .002]\). Post hoc tests showed that in this latency window, the ERP waveforms for the congruent audiovisual stimuli were more negative in the early phase than in the late phase. The LP condition evoked more negative waveforms than did the HP condition in the early phase.

In the early phase, the topographical map of the difference waveforms between HP and LP conditions during 90–120 ms interval showed a distribution in the frontal-central scalp areas (see Fig. 3). The measures of peak amplitude and latency demonstrated that this component reached its peak at about 102 ms and revealed the largest amplitude at the FCz electrode site \((1.4 \mu V)\).

3.2.2. Late phase: probability effects during the 40–96 ms interval

In the latency window of 40–96 ms, the results of the ANOVAs revealed significant main effects for probability condition \([F(1, 13) = 7.015, p < .05]\), as well as a significant phase × probability condition interaction effect \([F(1, 13) = 9.933, p < .01]\). A simple effects analysis found a significant main effect of probability in
the late phase [$F(1, 13) = 16.34, p < .001$], but not in the early phase [$F(1, 13) = 1.25, p = .283$]. A marginal main effect was observed for phase in both the HP condition [$F(1, 13) = 4.44, p = .055$] and the LP condition [$F(1, 13) = 4.27, p = .059$]. Post hoc tests showed that in this latency window, the ERP waveforms evoked by the HP condition were more negative in the late phase than in the early phase. The ERP waveforms evoked by the LP condition were more negative in the early phase than in the late phase. In the early phase, there was no difference between ERP waveforms evoked by HP and LP conditions.

In the late phase, the topographical map of the difference waveforms between the HP and the LP conditions during 40–96 ms interval demonstrate a distribution in the right parietal-occipital scalp areas (see Fig. 3). Measures of peak amplitude and latency demonstrated that this difference reached its peak at about 70 ms and revealed the largest amplitude at the PO4 electrode site (~1.3 μV). Due to the negative ERP waveforms, this difference suggests that congruent audiovisual stimuli elicited a more negative potential in the HP condition than in the LP condition in the late phase.

### 3.2.3. Late phase: probability effect during the 170–200 ms interval

In the latency window of 170–200 ms, the results of the ANOVA revealed a significant phase × probability condition interaction effect [$F(1, 13) = 14.056, p < .01$]. The positive ERP waveforms indicate that congruent audiovisual stimuli evoked a more positive potential in the frontal-central scalp areas in the HP condition than in the LP condition in the late phase. A simple effects analysis found a significant main effect of probability in the late phase [$F(1, 13) = 10.54, p < .01$], but not in the early phase [$F(1, 13) = 0.27, p > .05$]. A marginal main effect of phase was observed in the HP condition [$F(1, 13) = 3.5, p = .084$], but not in the LP condition [$F(1, 13) = 2.8, p > .05$]. Post hoc tests showed that in this latency window, the HP condition in the late phase resulted in more positive waveforms than the HP condition in the early phase, the LP condition in the early phase, or the LP condition in the late phase.

In the late phase, the scalp distribution of the difference waveforms between the HP and the LP conditions in the interval of 170–200 ms showed a frontal-central distribution (see Fig. 3). The measures of peak amplitude and latency demonstrated that this component reached its peak at about 188 ms and revealed the largest amplitude at the Fz electrode site (1.6 μV).

### 4. Discussion

This study contributes to our understanding of the neural bases underlying the effects of prior knowledge on audiovisual integration. This was achieved by analyzing the changes in both behavioral and ERP measures when the frequency of the audiovisual congruency was manipulated.

#### 4.1. Behavioral results

The results of the behavioral data showed that reaction times for congruent audiovisual stimuli in the HP and LP conditions were significant faster than that for incongruent audiovisual stimuli in the LP conditions. This means that the task-irrelevant auditory stimuli influenced the processing of visual letter identification. Moreover, for the congruent audiovisual stimuli, while the RT difference between the HP and the LP conditions during the first three blocks (blocks 1, 2 and 3) was not significant, RTs during blocks 4 and 5 were faster in the HP condition than in the LP condition. This indicates that participants were not able to distinguish the two probability conditions by means of letter color during the early phase of experiment. However, after experiencing three experimental blocks (900 trials), participants seemed to implicitly relate the probabilities of congruent audiovisual stimuli with letter color. This implicitly obtained knowledge then led to a difference in RTs in the HP and LP conditions for congruent audiovisual stimuli.

It has been suggested that cues are integrated into a single unified percept only if they are perceived to result from a common source rather than from different sources (Vatakis and Spence, 2008; Welch, 1999; Welch and Warren, 1980). Along with this assumption, a Bayesian causal inference model has also been proposed. According to this model, during the presentation of multisensory cues, the estimate for the cue from one sensory modality would be the sum of the multisensory statistically optimal estimate weighted by the probability of the common source and the corresponding unsensory estimate weighted by the probability of the different sources (Körding et al., 2007). The probability that the sensory cues result from a common source can be computed based on several factors including, the reliability of the individual sensory cues, the disparity of the sensory cues, and the prior probability that the sensory cues typically originate from a common source (see Ma and Pouget, 2008 for a review). Thus, a reduction in the “common source” probability would also lead to a reduction in multisensory integration. In line with this model, in the current experiment the slower response for the audiovisual congruent stimuli in the LP condition could be due to the frequent presentation of incongruent audiovisual stimuli. This may have resulted in the reduction of the common source prior probability, which then lead to the response using only visual cues in a much higher proportion of trials. Finally, due to the fact that after the completion of

### Table 1

<table>
<thead>
<tr>
<th>Probability</th>
<th>Time window</th>
<th>Block 1</th>
<th>Block 2</th>
<th>Block 3</th>
<th>Block 4</th>
<th>Block 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>HP congruency</td>
<td></td>
<td>96.4 ± 3.3%</td>
<td>96.6 ± 2.8%</td>
<td>96.1 ± 2.9%</td>
<td>96.4 ± 2.7%</td>
<td>95.1 ± 3.9%</td>
</tr>
<tr>
<td>LP congruency</td>
<td></td>
<td>95.0 ± 6.9%</td>
<td>94.2 ± 6.1%</td>
<td>94.9 ± 5.2%</td>
<td>94.5 ± 6.0%</td>
<td>94.0 ± 8.7%</td>
</tr>
<tr>
<td>LP incongruency</td>
<td></td>
<td>91.0 ± 9.6%</td>
<td>92.9 ± 6.5%</td>
<td>93.5 ± 6.9%</td>
<td>93.5 ± 7.1%</td>
<td>94.4 ± 6.3%</td>
</tr>
</tbody>
</table>

### Table 2

<table>
<thead>
<tr>
<th>Probability</th>
<th>Time window</th>
<th>Block 1</th>
<th>Block 2</th>
<th>Block 3</th>
<th>Block 4</th>
<th>Block 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>HP congruency</td>
<td></td>
<td>448 ± 49</td>
<td>432 ± 41</td>
<td>439 ± 39</td>
<td>433 ± 36</td>
<td>432 ± 43</td>
</tr>
<tr>
<td>LP congruency</td>
<td></td>
<td>451 ± 49</td>
<td>441 ± 43</td>
<td>442 ± 47</td>
<td>445 ± 44</td>
<td>445 ± 41</td>
</tr>
<tr>
<td>LP incongruency</td>
<td></td>
<td>491 ± 56</td>
<td>475 ± 45</td>
<td>470 ± 45</td>
<td>461 ± 44</td>
<td>463 ± 44</td>
</tr>
</tbody>
</table>
Fig. 1. The grand average ERPs for the congruent audiovisual stimuli for the AV$_{HP}$ (dashed lines) and AV$_{LP}$ (gray lines) conditions and the AV$_{HP}$ – AV$_{LP}$ difference waveforms (black lines) at the frontal (F3, Fz, F4), central (C3, Cz, C4), and parietal (P3, Pz, P4) electrode sites.

The study participants indicated that they were unaware that cue congruency was contingent on the color of the letter, this suggests that the acquisition of this prior knowledge is likely an implicit process.

4.2. ERP results

As described above, based on the obtained results from the behavioral task, the first 600 trials (blocks 1 and 2) were catego-
Fig. 2. Statistical significance of the residual AVHP−AVLP activities between 0 and 300 ms for all participants (Student’s t-tests comparing the amplitude of the AVHP−AVLP difference wave against zero at each latency). a. Effects at the early phase. b. Effects at the late phase.

This result indicates that the two phase category designations used for analyzing the ERP waveforms were able to effectively distinguish between the different stages of obtaining prior knowledge and how this might then affect multisensory integration.

4.2.1. ERP in the early phase of the experiment: probability effects during the 90–120 ms interval

In the interval of 90–120 ms, a significant probability effect was found over the frontal-central scalp areas in the early phase. The scalp distribution of the effect is similar to the audiovisual integration effects at around 100 ms described by previous studies (Fort et al., 2002; Molholm et al., 2002; Teder-Sälejärvi et al., 2002; Giard and Peronnet, 1999; van Wassenhove et al., 2005; Besle et al., 2004; Klucharev et al., 2003; Stekelenburg and Vroomen, 2007). Furthermore, it has been reported that in frontal regions, audiovisual integration effects at around 100 ms were larger for attended audiovisual stimuli than for unattended audiovisual stimuli (Talsma and Woldorff, 2005).

Therefore, one likely explanation for the current results is that attentional effects may have modulated audiovisual integration. In particular, this may indicate that the brain assigns more attention to the audiovisual stimulus in the LP condition. Because, in the practice phase of the experiment, the audiovisual stimuli were always congruent, participants likely formed a common source prior for the auditory and visual letter cues. When the incongruent audiovisual stimuli were presented in the LP condition, participants may have implicitly detected that the probability of cue congruency was no longer in accord with the common source prior acquired during practice. It would, therefore, be reasonable to assume that the brain might attend more to the colored letter with the lower congruency so as to better establish a new prior through repetitive learning and thus increase task performance. Therefore, the effect observed here is likely related to the amount of attention assigned to accumulate prior knowledge.

For our current results, an alternative, yet less likely, explanation for the ERP findings is that they reflect a modulating effect of prior knowledge on audiovisual integration. This explanation is perhaps less likely because the behavioral results indicate that participants were not yet able to associate the probability of congruency with color information in the early phase. In addition, if the effect was indeed correlated with the modulating effect of prior knowledge on audiovisual integration, then a similar or even stronger modulation should have been observed during the late phase, which it was not.

4.2.2. ERP in the late phase of the experiment: probability effects during the 40–96 ms interval

The current ERP data showed that an early probability effect during the 40–96 ms interval only occurred in the late phase of the experiment with the topographical map of the difference wave-
forms showing a distribution in the right parietal-occipital scalp areas. Previous ERP studies investigating audiovisual integration have reported an early audiovisual interaction arising over visual cortex for bimodal audiovisual stimuli, which has a similar scalp distribution and latency to the current early probability effect (Giard and Peronnet, 1999; Molholm et al., 2002; Klucharev et al., 2003). Thus, the current early probability effect could be related to audiovisual interaction. In addition to audiovisual integration research, early multisensory interactions in sensory-specific cortices have also been observed in studies of audio-somatosensory integration (Foxe et al., 2000; Murray et al., 2005). Molholm et al. (2002) suggest that the early audiovisual interaction found in their study could result from an enhancement of attention for processing visual stimuli which was exogenous driven by temporally coincident auditory stimuli.

Thus, a possible explanation for our results is that the early probability effect resulted from a modulation of prior knowledge on visual exogenous attention which was driven by temporally coincident auditory stimuli. Furthermore, the early probability effect may result from the fact that the congruent audiovisual stimuli elicited a more negative potential in the HP condition than in the LP condition. This means that, for the HP condition, the task-irrelevant auditory stimulus, which was congruent with visual stimulus, would result in a greater enhancement of attention for processing visual stimuli than for the LP condition. Moreover, an alternative explanation for the ERP findings could be that the early probability effect reflects different expressions of prior knowledge in the different experimental stages (i.e. the strength of neural activity indicates that cue-congruency probabilities were detected). This assumption suggests that prior knowledge could have been stored in the sensory-specific cortex, extracted on the basis of some attributes of an object or event, and then transmitted together with sensory-specific information to multisensory cortices where cross-modal integration occurred. Although both hypotheses remain to be confirmed by further studies, the extraction of prior knowledge most likely occurs at early stages of perceptual processing regardless of which explanation is true.

4.2.3. ERP in the late phase of the experiment: probability effect during the 170–200 ms interval

In the current study, the third probability effect was found during the 170–200 ms interval in the late phase. This effect had a frontal-central distribution with similar latency and scalp distributions to that of classic auditory P2 and may thus reflect increased activity of the auditory P2 generators. Previous studies have shown an enhanced P2 amplitude in response to bimodal audiovisual information (Spreckelmeyer et al., 2006; Giard and Peronnet, 1999; Molholm et al., 2002; Stekelenburg and Vroomen, 2007; Vidal et al., 2008). Auditory P2 could be a dipolar vertex-positive auditory-evoked potential generated by cortical sources lateral to superior temporal gyrus (Scherg et al., 1989; Picton et al., 1999; Pantev et al., 1996). Further, neuroimaging studies in humans have shown that the superior temporal gyrus plays a significant role during the processing of audiovisual integration for lip movements and speech sounds (Calvert et al., 1999, 2000; Sams et al., 1991; Sekiyama et al., 2003; Paulesu et al., 2003; Callan et al., 2001; Wright et al., 2003; Reale et al., 2007), visual letter and speech sounds (van Atteveldt et al., 2004, 2007; Raji et al., 2000), visual flash and auditory tones (Teder-Sälejärvi et al., 2005), and visual animal picture and animal sound pairs (Beauchamp et al., 2004; Hein et al., 2007). Single-cell recordings in monkeys have also shown an activation of the superior temporal gyrus for audiovisual stimuli (Bruce et al., 1981; Hikosaka, 1993; Schroeder and Foxe, 2002; Benevento et al., 1977). These lines of evidence suggest that this P2 modulation was associated with the activation of multisensory cortical areas during audiovisual integration. Therefore, the probability effect during the 170–200 ms interval represents a modulating effect of audiovisual interactions by prior knowledge of cue-congruency probabilities. The more positive ERPs evoked in the HP condition compared to the LP condition could be interpreted as a stronger neural activation in areas involved in visual-auditory information processing. The results provide support for the hypothesis that prior knowledge specifying a high probability of cue-congruency improves audiovisual integration compared to a low probability of cue-congruency.

4.3. Related findings and implications of our results

Recently there have been several ERP studies investigating the phenomena of cross-modal spread of attention using an experimental task similar to the task used in the current experiment, in which participants were asked to attend only to the visual stimuli and ignore the auditory stimuli (Fiebelkorn et al., 2010; Busse et al., 2005; Zimmer et al., 2010). For example, in a study by Busse et al. (2005), the effect of the spread of attention from the visual stimulus (task-relevant) to the auditory stimulus (task-irrelevant and from a different location) was revealed through the comparison of the ERP waveform elicited by the auditory stimuli (audiovisual combined waveform minus visual only waveform) that occurred with an attended visual stimulus, relative to that occurred with an unattended one. In the Busse et al. (2005) study, the visual stimulus (a group of dots) and the auditory stimulus (tone) were not inherently related. In Zimmer et al. (2010), however, the visual stimulus (a letter) was accompanied by either a congruent or an incongruent auditory stimulus (spoken letter that was the same or different than the viewed letter). They also found the same ERP pattern for the spread of attention. Moreover, the incongruent auditory stimulus elicited even larger ERP waveforms for the spread of attention, reflecting the conflict detection and multisensory interaction processes at a semantic processing stage.

In the early trials of our experiment, our test was somewhat similar to what was implemented by Zimmer et al. (2010) in that both congruent and incongruent letter–sound pairs were included except that in Zimmer et al. (2010), no additional information was provided about the probability of congruency. In our study, because we provided a color cue as an implicit indication of the probability of congruency, for the same congruent letter–sound pairs, we found an ERP waveform difference when the color of the letters indicated a high probability relative to a low probability of congruency. Although our paradigm could not offer direct evidence for the effect of attention, our results suggest that different extent of attention have been allocated for the same congruent letter–sound pairs when the exposure of different probabilities of the congruency is accumulated in the early phase of the experiment.

Therefore our paradigm offers a unique opportunity to reveal the underlying processes in which the knowledge about the color-congruency association is acquired. We discovered distinct ERP waveforms in the early phase of the experiment (knowledge acquisition phase) and the late phase of the experiment (knowledge utilization phase), indicated by difference of RTs for congruent pairs between the HP and the LP conditions. Our study represents the first study to directly examine brain activities related to the prior knowledge for multisensory integration.

The behavioral and ERP findings generated from our study have important implications for the role of the “unity assumption” in interpreting the process of multisensory integration. In most multisensory integration studies, it is often assumed that sensory cues from different modalities come from the same source (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). However, in the real world, information from the environment can be generated from the same or different sources, with the information about the source poten-
tially informed by contextual cues. The role of the prior has only recently been experimentally studied by a few labs using behavioral tasks (e.g., Atkins et al., 2001). We hope our new electrophysiologically findings about the time course of the underlying process during and following the acquisition of the prior will trigger more research on this very important topic.

5. Conclusions

The present findings provided through transient ERP recordings have provided insights into the neural bases of the effects of prior knowledge on audiovisual integration. The behavioral results indicate that human observers gradually and implicitly update prior knowledge regarding the probability of congruent audiovisual stimuli. This knowledge is then used to modify audiovisual integration in a context-dependent manner. The ERP results indicate that, during the period in which new prior knowledge was being acquired (i.e., the early phase), the brain attended more to the audiovisual stimuli whose probability of cue congruency changed frequently. This then allowed for a more rapid acquisition of important information related to expected frequency of cue congruences. Moreover, during the process of auditory–visual integration, prior knowledge was extracted at early stages of visual processing and subsequently modulated activity in multisensory cortical areas.

Acknowledgements

This study was supported by awards from the Postgraduate Innovation Foundation of Science and Technology at Southwest University (h2007004) and the National Key Discipline of Basic Psychology at Southwest University (NSKD08013) and the Fundamental Research Funds for the Central Universities (SWU1009009) and the Natural Science and Engineering Research Council of Canada.

References


Hikosaka, K., 1993. The polysensory region in the anterior bank of the caudal superior temporal sulcus of the macaque monkey. Biomedical Research (Tokyo) 14, 1–45.


