

Dynamic characteristics of multisensory facilitation and inhibition

W. Y. Wang · L. Hu · E. Valentini ·
X. B. Xie · H. Y. Cui · Y. Hu

Received: 29 April 2011 / Revised: 14 February 2012 / Accepted: 1 March 2012 / Published online: 20 March 2012
© Springer Science+Business Media B.V. 2012

Abstract Multimodal integration, which mainly refers to multisensory facilitation and multisensory inhibition, is the process of merging multisensory information in the human brain. However, the neural mechanisms underlying the dynamic characteristics of multimodal integration are not fully understood. The objective of this study is to investigate the basic mechanisms of multimodal integration by assessing the intermodal influences of vision, audition, and somatosensory sensations (the influence of multisensory background events to the target event). We used a timed target detection task, and measured both behavioral and electroencephalographic responses to visual target events (green solid circle), auditory target events (2 kHz pure tone) and somatosensory target events (1.5 ± 0.1 mA square wave pulse) from 20 normal participants. There were significant differences in both behavior performance

and ERP components when comparing the unimodal target stimuli with multimodal (bimodal and trimodal) target stimuli for all target groups. Significant correlation among reaction time and P3 latency was observed across all target conditions. The perceptual processing of auditory target events (A) was inhibited by the background events, while the perceptual processing of somatosensory target events (S) was facilitated by the background events. In contrast, the perceptual processing of visual target events (V) remained impervious to multisensory background events.

Keywords Multisensory enhancement · Event-related potentials · P3 latency · Visual dominance · Source analysis

Background

It is common for the human brain to process multiple types of sensory information simultaneously. This is called multimodal perception (Arabzadeh et al. 2008; Driver and Noesselt 2008). Multimodal facilitation has been reported in several articles (Arabzadeh et al. 2008; Bresciani et al. 2008; Chen and Yeh 2009), where the authors emphasized that the processing of stimuli belonging to different sensory modalities can be facilitated by the simultaneous processing of a unimodal stimulus (Driver and Spence 1998; Meredith and Stein 1986; Teder-Salejarvi et al. 2005). Bresciani et al. (2008) investigated the interaction of visual, auditory, and tactile sensory information during the presentation of sequences of events, and found that while vision, audition, and touch information were automatically integrated, their respective contributions to the integrated percept were different. In a bimodal setting, Chen and Yeh (2009) showed that the perception of visual events was enhanced by accompaniment with auditory stimulation,

W. Y. Wang and L. Hu contributed equally to this work.

W. Y. Wang · X. B. Xie · H. Y. Cui · Y. Hu
Chinese Academy of Medical Science and Peking Union
Medical College, Institute of Biomedical Engineering, Tianjin,
China

W. Y. Wang · Y. Hu (✉)
Department of Orthopaedics and Traumatology, Duchess of Kent
Children's Hospital, The University of Hong Kong, 12 Sandy
Bay Road, Hong Kong, China
e-mail: yhud@hku.hk

L. Hu
Key Laboratory of Cognition and Personality (Ministry of
Education), School of Psychology, Southwest University,
Chongqing, China

E. Valentini
Department of Neuroscience, Physiology and Pharmacology,
University College London, London, UK

thus supporting the facilitation effect of sound on vision. Similarly, Odgaard et al. (2004) showed that concurrent visual stimulation could enhance the loudness of auditory white noise, and Arabzadeh et al. (2008) confirmed that a visual stimulation could improve the discrimination of somatosensory stimulation, thus providing support for the facilitation effect of vision on both auditory and somatosensory modalities. Some researchers have also analyzed multisensory facilitation using electrophysiological methods at both the cortical level and neuronal level (Jiang et al. 2002; McDonald et al. 2003; Molholm et al. 2002; Rowland et al. 2007).

Conversely, some studies have stressed the importance of an inhibition effect in multimodal perception (e.g. the Colavita visual dominance effect) (Colavita 1974). Indeed, simultaneous presentation of multiple stimuli can contribute to a decrease in the ability to perceive or respond to unimodal stimuli (Hartcher-O'Brien et al. 2008; Hecht and Reiner 2009; Koppen and Spence 2007; Occelli et al. 2010). The Colavita visual dominance effect has repeatedly been reported with vision and audition (Koppen et al. 2009; Van Damme et al. 2009), and with vision and somatosensation (Hecht and Reiner 2009). This effect has been attributed to the higher dominance of attentional and arousal resources related to vision over resources required for processing audition and somatosensation.

Notwithstanding the bulk of research on multisensory interactions, the influence of simultaneous multisensory background events on the processing of target events (facilitation, inhibition, or both) is unclear. Sinnett et al. (2008) observed that both multisensory facilitation and inhibition could be observed in the reaction to the same bimodal event, leading them to propose the co-occurrence of these processes. However, the neural mechanisms subserving the co-occurrence of multisensory facilitation and inhibition have not been explicitly investigated.

In this study, we used a timed target detection task to investigate multisensory integration and its underlying neural mechanisms by (1) comparing behavioral responses (reaction times [RTs]) in unimodal (visual, auditory, and somatosensory), bimodal (visual-auditory, visual-somatosensory, and auditory-somatosensory), and trimodal (visual-auditory-somatosensory) experimental conditions; and (2) analyzing ERP responses (the peak latency and amplitude of N1, P2, and P3).

Methods

Participants

Twenty right-handed healthy volunteers (twelve females and eight males who were undergraduate and graduate

students from Tianjin University), aged 20 to 26 years (22.5 ± 0.2 , mean $\pm$ SD), took part in the experiment. All of the participants reported normal hearing, normal sensorimotor function, and normal or corrected-to-normal vision. All participants gave written informed consent, and the local ethics committee approved the procedures.

Stimuli

The visual stimuli, including the target event (V) (green solid circle: diameter = 3 cm; angle of view = 2.29° ; duration = 60 ms) and non-target event (v) (red solid circle: diameter = 3 cm; angle of view = 2.29° ; duration = 60 ms) (Molholm et al. 2002; Feng et al. 2008), consisted of flashes presented in the center of a 17 in. CRT monitor placed in front of the participant, at a distance of 75 cm (Fig. 1, top panel, left). The auditory stimuli, including the target event (A) (high frequency monotone: frequency = 2 kHz; SPL = 80 dB; duration = 60 ms, with 5 ms rise/fall times) and non-target event (a) (low frequency monotone: frequency = 1 kHz; SPL = 80 dB; duration = 60 ms, with 5 ms rise/fall times) (Molholm et al. 2002; Feng et al. 2008), were sound bursts delivered to the participant through binaural earphones (Stim Audio System, NeuroScan Lab) (Fig. 1, top panel, center). The somatosensory stimuli, including the target event (S) (higher intensity: stimuli intensity = 1.5 ± 0.1 mA [three times the sensory threshold]; duration = 1 ms) and non-target event (s) (lower intensity: stimuli intensity = 1.0 ± 0.1 mA [twice the sensory threshold]; duration = 1 ms) (Tanaka et al. 2008), consisted of square wave pulses administered transcutaneously to the left index finger via metallic rings (Fig. 1, top panel, right). The presentation time of the visual, auditory, and somatosensory stimuli was synchronized using a computer with STIM software (NeuroScan Lab).

Procedure

Participants were seated in a comfortable chair in a dimly lit electrically shielded room, and were asked to focus their attention on the occurrence of the stimuli, regardless of modality. In the experiment, unimodal (visual, auditory, and somatosensory), bimodal (visual-auditory, visual-somatosensory, and auditory-somatosensory), and trimodal (visual-auditory-somatosensory) stimuli were presented in random order (Fig. 1, bottom panel). The participants were instructed to monitor all events (visual, auditory, and somatosensory), regardless of their status as target or non-target, and to respond as quickly as possible to the appearance of a predefined target by pressing a response button (STIM SYSTEM SWITCH RESPONSE PAD,

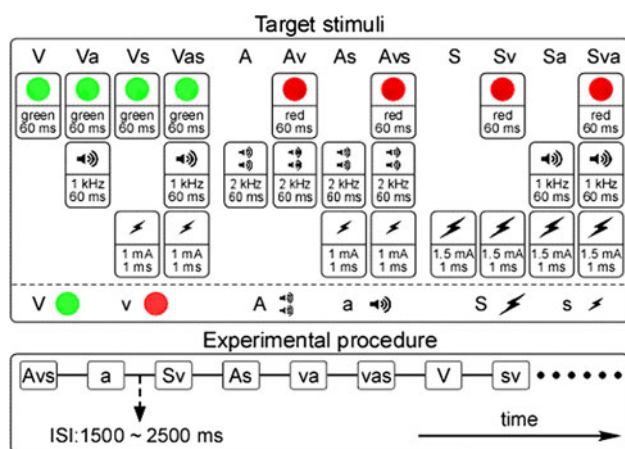


Fig. 1 The illustration of target stimuli and experimental procedure. *Top panel:* Twelve target stimuli (i.e. visual target group: V, Va, Vs, Vas; auditory target group: A, Av, As, Avs; and somatosensory target group: S, Sv, Sa, Sva). One column represents one target stimuli. For visual stimuli, the target event (V) is a green flash (solid circle; duration = 60 ms), whereas the non-target event (v) is a red flash (solid circle; duration = 60 ms). For auditory stimuli, the target event (A) is a high frequency monotone sound burst (frequency = 2 kHz; duration = 60 ms), whereas the non-target event (a) is a low frequency monotone sound burst (frequency = 1 kHz; duration = 60 ms). For somatosensory stimuli, the target event (S) is a higher intensity electrical stimulus (stimuli intensity = 1.5 ± 0.1 mA; duration = 1 ms), whereas the non-target event (s) is a lower intensity electrical stimulus (stimuli intensity = 1.0 ± 0.1 mA; duration = 1 ms). *Bottom panel:* Both target stimuli and non-target stimuli were presented in random order with the ISI varying randomly between 1,500 and 2,500 ms

NeuroScan Lab), with their right index finger. The target and non-target stimuli were randomly presented (target stimuli: 50 %; non-target stimuli: 50 %).

For the non-target stimuli, there were seven trial types, which were combined by the unimodal, bimodal, and trimodal stimuli (i.e. v, a, s, va, vs, as, vas, with 144 trials for each non-target trial type). Similarly, for the target stimuli, to which participants had to respond, there were twelve trial types with four in each of the three target groups (visual target group: V, Va, Vs, Vas; auditory target group: A, Av, As, Avs; and somatosensory target group: S, Sv, Sa, Sva) (Fig. 1, top panel). For each target stimulus event, 72 stimuli were delivered to the participant during the experiment. The inter-stimulus interval (ISI) varied randomly between 1,500 and 2,500 ms (Fig 1, bottom panel). Prior to data collection, the participants were given 200 practice trials, which included all conditions, to familiarize them with the task. Altogether, the experimental session was divided into six blocks that were each about 13 min long, with a 5 min to 10 min break between successive blocks, and the entire experiment took less than 120 min for each participant.

Behavioral evaluation and analysis

The RTs that were shorter than 200 ms and longer than 1,000 ms were considered to be anticipations and misses, respectively, and were not included in the analyses. RTs among the unimodal, bimodal and trimodal conditions were compared using a one-way repeated-measures analysis of variance (ANOVA) with a statistical significance level of $P < 0.05$. Following this, post hoc comparisons were conducted using paired t-tests.

EEG data recording and analysis

The EEG data were recorded using a 64-channel NeuroScan system (pass band: 0.05–100 Hz, sampling rate: 1,000 Hz) and a standard EEG cap based on the extended 10–20 system of electrode placement. Cz was used as the reference channel, and all channel impedances were kept under 5 k Ω . To monitor ocular movements and eye blinks, electro-oculographic signals were simultaneously recorded from four surface electrodes, with one pair placed over the upper and lower eyelid, and the other pair placed 1 cm lateral to the outer corner of the left and right orbit.

EEG data were processed using Scan (NeuroScan Lab) and EEGLAB, an open source toolbox running under the MATLAB environment. Continuous EEG data were band-pass filtered between 0.1 and 30 Hz. EEG epochs, which were segmented in 1,200 ms time-windows (200 ms pre-stimulus and 1,000 ms post-stimulus). The baseline was corrected using the pre-stimulus time interval. Trials contaminated by eye-blinks and movements were removed using the automated ocular artifact reduction module in Scan (NeuroScan Lab). After artifact rejection and baseline correction, EEG trials were re-referenced to the bilateral mastoid electrodes (Schinkel et al. 2007). For each participant and each stimulus condition, the average waveforms were computed and time-locked to the onset of the stimulus. Single-participant average waveforms were subsequently averaged to obtain group-level average waveforms for each experimental condition.

Using the average ERPs for each participant, we measured the peak latency and baseline-to-peak amplitude (with respect to the mean amplitude in the baseline time interval, the same hereinafter) of N1 (between 120 and 180 ms; at P5 and P6 [(P5 + P6)/2]), P2 (between 180 and 280 ms; at Cz), and P3 (between 280 and 600 ms; at Pz) for the visual target group (Luck et al. 2005; Beim Graben et al. 2008), the peak latency and baseline-to-peak amplitude of N1 (between 100 and 160 ms; at Fz), P2 (between 180 and 280 ms; at Cz), and P3 (between 300 and 650 ms; at Pz) for auditory target group (Luck et al. 2005; Guntekin and Basar 2010), and the peak latency and baseline-to-peak

amplitude of N1 (between 100 and 180 ms; at C4), P2 (between 180 and 280 ms; at Cz), and P3 (between 300 and 650 ms; at Pz) for the somatosensory target group (Luck et al. 2005). For each condition, the group-level scalp topographies of N1, P2, and P3 components were computed using spline interpolation. For the three target groups, a one-way repeated-measures ANOVA was used to compare the peak latencies and amplitudes. When significant, we performed post hoc analyses using the Tukey correction to compare the latencies and amplitudes of the different target groups. Such statistical comparison (among peak parameters of target ERPs) that was frequently used in previous studies (Sinnett et al. 2008; Feng et al. 2008), was able to assess the intermodal influences of multisensory background stimuli on the target ERPs, which could help investigating the basic mechanisms of multimodal integration in the present study.

Correlation analysis

The measured N1, P2, and P3 latencies, as well as the N1, P2, and P3 amplitudes, were correlated with the corresponding mean RTs using Linear Regression analysis. For each condition, latencies and amplitudes of N1, P2, and P3 were averaged across participants and compared with the corresponding RTs, which were also averaged across participants. Correlation coefficients and their significance levels were calculated for each of these values.

Results

Behavioral results

There was a significant difference in behavioral performance (RT) among the unimodal (V, S, A), bimodal (either 2 modalities of V, S, A) and trimodal (all 3 modalities) conditions ($F = 17.163$, $P < 0.001$). As shown in Fig. 2, there was a significant multimodal effect of sensory stimuli on auditory and somatosensory targets. In contrast, there was no significant multimodal effect of sensory stimuli on visual targets.

RTs to auditory stimuli were significantly shorter in the unimodal condition (A) than in the bimodal condition (Av, As) ($A < Av$, $P = 0.025$; $A < As$, $P = 0.004$) or trimodal condition ($A < Asv$, $P < 0.001$). RTs to auditory stimuli were not significantly different between the bimodal condition (Av, As) and the trimodal condition (Avs) (Av vs. Avs , $P = 0.966$; As vs. Avs , $P = 0.505$).

RTs to somatosensory stimuli were significantly longer in the unimodal condition (S) than the bimodal condition (Sv, Sa) and the trimodal condition (Sva) ($S > Sa$, $P = 0.029$; $S > Sv$, $P < 0.001$; $S > Sva$, $P < 0.001$).

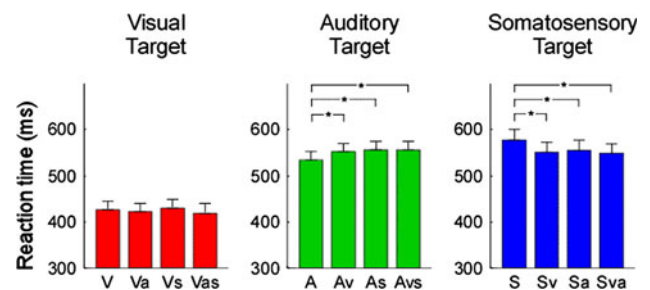


Fig. 2 Comparison of reaction times (RTs). The RTs were displayed in red, green, and blue for visual, auditory, and somatosensory target groups respectively. Error bars represent, for each condition, $\pm$ SE across subjects. Asterisk indicates a significant difference ($P < 0.05$, two tailed paired t test). Noteworthy is the general faster reaction time during vision than during somatosensory or auditory stimulation. Also note, the opposite pattern in auditory (faster response during uni-modal condition) and tactile (faster response during bi-modal and tri-modal) modality

In contrast with the auditory and somatosensory responses, the RT in the visual target group was not significantly different among the unimodal (V), bimodal (Va, Vs) or trimodal conditions (Vas) (V vs. Va , $P = 0.457$; V vs. Vs , $P = 0.464$; V vs. Vas , $P = 0.369$).

ERP results

The latencies and amplitudes of N1, P2 and P3 were significantly different among the unimodal (V, S, A), bimodal (either 2 modalities of V, S, A) and trimodal (all 3 modalities) conditions (N1 latency: $F = 3.057$, $P < 0.001$; N1 amplitude: $F = 18.78$, $P < 0.001$; P2 latency: $F = 4.114$, $P < 0.001$; P2 amplitude: $F = 5.6$, $P < 0.001$; P3 latency: $F = 4.668$, $P < 0.001$; P3 amplitude: $F = 2.154$, $P = 0.018$, one-way ANOVA). Post hoc analysis revealed a significant cross-modality multimodal effect among visual, auditory and somatosensory.

Visual target

Group-level average waveforms and scalp topographies of N1, P2, and P3 for unimodal (V), bimodal (Va, Vs) and trimodal (Vas) conditions in the visual target group were shown in Fig. 3. While the scalp topographies of N1 were largely different among conditions (negative activity was maximal at the frontal region when an auditory non-target event occurred), the scalp topographies of both P2 (positive activity was maximal at the occipital region) and P3 (positive activity was maximal at the parietal region) were similar across conditions.

Statistical comparisons of ERP peak parameters (latency and amplitude) in terms of N1, P2, and P3 were performed using a one-way ANOVA (Table 1). Significant differences were found in N1 amplitude ($F = 3.361$,

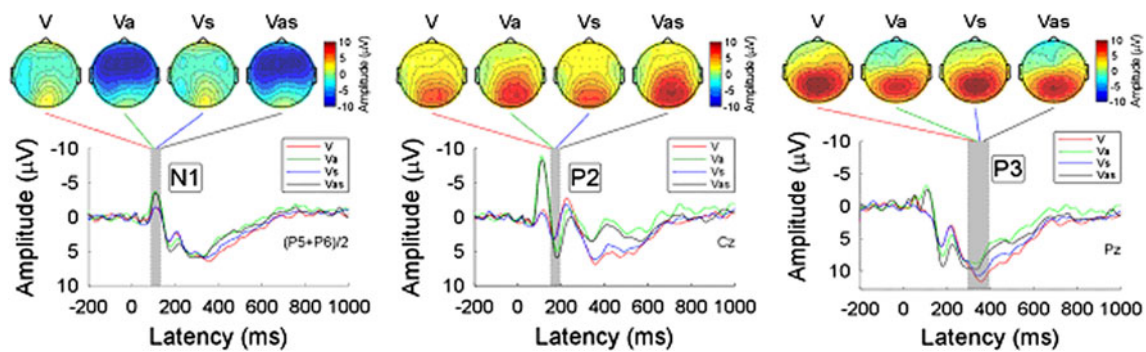


Fig. 3 Group-level average waveforms and scalp topographies of N1, P2, and P3 at four conditions (V, Va, Vs, and Vas) within visual target group. Group-level average waveforms at V, Va, Vs, and Vas conditions are presented using red, green, blue, and black lines. The gray region in each panel shows the time interval to measure ERP peaks. *Left panel:* Group-level average waveforms (measured at

[P5 + P6]/2) and scalp topographies of N1 at four conditions. *Middle panel:* Group-level average waveforms (measured at Cz) and scalp topographies of P2 at four conditions. *Right panel:* Group-level average waveforms (measured at Pz) and scalp topographies of P3 at four conditions

$P = 0.023$), P2 latency ($F = 2.937$, $P = 0.039$), and P2 amplitude ($F = 3.567$, $P = 0.018$). There was no significant difference in N1 latency ($F = 0.362$, $P = 0.781$), P3 latency ($F = 0.681$, $P = 0.565$) and P3 amplitude ($F = 0.93$, $P = 0.43$).

Auditory target

Group-level average waveforms and scalp topographies of N1, P2, and P3 for the unimodal, bimodal and trimodal conditions in the auditory target group were shown in Fig. 4. While the scalp topographies of P2 were largely different among the conditions (positive activity was maximal at the central region and occipital region with and without a co-occurring visual non-target event respectively), the scalp topographies of both N1 (a negative maximum at the frontal region) and P3 (a positive maximum at the parietal region) were similar across the four conditions.

Statistical analysis using a one-way ANOVA revealed significant differences in N1 amplitude ($F = 2.952$, $P = 0.038$) and P2 latency ($F = 6.196$, $P < 0.001$), but no significant difference in N1 latency ($F = 0.2$, $P = 0.90$), P2 amplitude ($F = 0.934$, $P = 0.429$), P3 latency ($F = 0.019$, $P = 0.99$), and P3 amplitude ($F = 0.403$, $P = 0.75$) (Table 1).

Somatosensory target

Group-level average waveforms and scalp topographies of N1, P2, and P3 for the four conditions (S, Sv, Sa, and Sva) in the somatosensory target group were shown in Fig. 5. While the scalp topographies of both N1 and P2 were largely different among the conditions (N1 scalp topographies: negative maximum at the contralateral frontal region with a co-occurring auditory non-target event; P2 scalp

topographies: positive maximum at the central region and occipital region with and without a co-occurring visual non-target event respectively), the scalp topographies of P3 (a positive maximum at the parietal region) were similar across the four conditions.

Statistical analysis using a one-way ANOVA revealed significant differences in N1 latency ($F = 4.297$, $P = 0.007$), N1 amplitude ($F = 17.33$, $P < 0.001$), P2 latency ($F = 3.87$, $P = 0.012$), P2 amplitude ($F = 5.607$, $P = 0.002$), P3 latency ($F = 3.703$, $P = 0.015$), and P3 amplitude ($F = 3.035$, $P = 0.034$) (Table 1).

Correlation results

When correlating the behavioral data (reaction time) and ERP data (peak latency of N1, P2 and P3, and peak amplitude of N1, P2, and P3) for all conditions, we found significant correlations between RT and P3 latency ($R = 0.661$, $P < 0.001$) (Fig. 6).

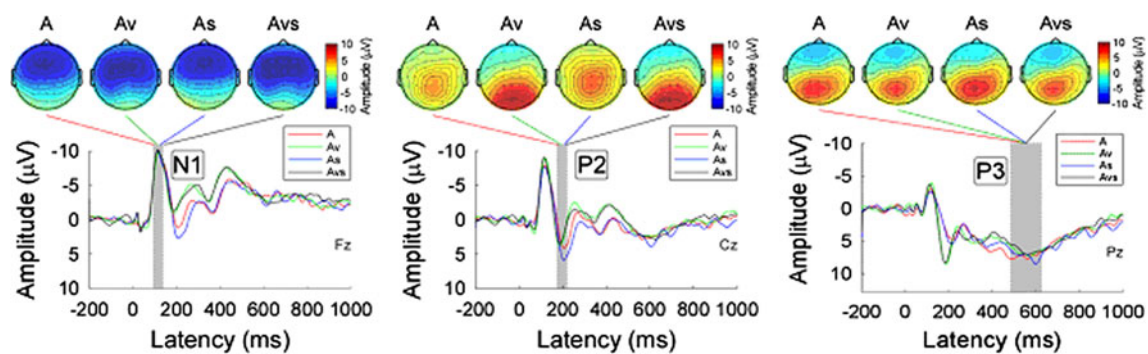
Discussion

In the present study, we adopted a timed target detection task to investigate multisensory interactions (both facilitation and inhibition) among vision, audition, and somatosensation. We obtained two main findings. First, non-target background events (v, s, and vs) inhibited the perceptual processing of auditory targets. Second, non-target background events (v, a, and va) enhanced the perceptual processing of somatosensory targets.

Our behavioral results suggested that, in general, target detection was faster during visual stimulation than during somatosensory or auditory stimulation. In addition, no differences between unimodal and multimodal trials were found for the visual target group. Slower responses were

Table 1 Behavioral results (reaction time), ERP results (the peak latency and amplitude of N1, P2, and P3), and their statistical comparison (one-way ANOVA)

Target sensory	Unimodal target	Multimodal target			One-way ANOVA	
	Mean $\pm$ SE	Bimodal A Mean $\pm$ SE	Bimodal B Mean $\pm$ SE	Trimodal Mean $\pm$ SE	F value	P value
RT (ms)						
V	426 $\pm$ 13	422 $\pm$ 13	429 $\pm$ 13	421 $\pm$ 15		
A	538 $\pm$ 14	555 $\pm$ 18	552 $\pm$ 14	555 $\pm$ 17		
S	576 $\pm$ 18	548 $\pm$ 17	552 $\pm$ 17	545 $\pm$ 16		
N1 lat. (ms)						
V	140 $\pm$ 4	139 $\pm$ 4	136 $\pm$ 3	136 $\pm$ 3	0.362	0.781
A	129 $\pm$ 4	130 $\pm$ 4	127 $\pm$ 4	126 $\pm$ 4	0.2	0.9
S	142 $\pm$ 6	125 $\pm$ 5	124 $\pm$ 5	118 $\pm$ 4	4.297	0.007
N1 amp. (μV)						
V	-3.4 $\pm$ 0.6	-5.4 $\pm$ 0.7	-3.6 $\pm$ 0.6	-5.5 $\pm$ 0.6	3.361	0.023
A	-10.7 $\pm$ 1	-10.9 $\pm$ 0.8	-11.3 $\pm$ 1	-11.7 $\pm$ 1	2.952	0.038
S	-2 $\pm$ 0.6	-8 $\pm$ 1	-3.8 $\pm$ 0.7	-10.1 $\pm$ 1.1	17.33	<0.001
P2 lat. (ms)						
V	193 $\pm$ 5	203 $\pm$ 4	195 $\pm$ 5	201 $\pm$ 4	2.937	0.039
A	220 $\pm$ 4	224 $\pm$ 6	199 $\pm$ 5	201 $\pm$ 6	6.196	<0.001
S	215 $\pm$ 8	210 $\pm$ 5	190 $\pm$ 8	192 $\pm$ 5	3.87	0.012
P2 amp. (μV)						
V	3.6 $\pm$ 1.1	6.6 $\pm$ 1.3	4.5 $\pm$ 1	7.3 $\pm$ 1.1	3.567	0.018
A	5.1 $\pm$ 1.1	6.9 $\pm$ 1.2	4.6 $\pm$ 1.1	4.8 $\pm$ 1.1	0.934	0.429
S	7.1 $\pm$ 0.8	12.6 $\pm$ 1.4	6.9 $\pm$ 1	11.6 $\pm$ 1.7	5.607	0.002
P3 lat. (ms)						
V	403 $\pm$ 12	392 $\pm$ 11	394 $\pm$ 11	381 $\pm$ 10	0.681	0.565
A	455 $\pm$ 17	467 $\pm$ 21	470 $\pm$ 21	470 $\pm$ 20	0.019	0.99
S	479 $\pm$ 18	452 $\pm$ 16	455 $\pm$ 17	449 $\pm$ 19	3.703	0.015
P3 amp. (μV)						
V	13.3 $\pm$ 1.7	9.9 $\pm$ 1.5	12.4 $\pm$ 1.6	10.3 $\pm$ 1.4	0.93	0.43
A	9.4 $\pm$ 1.5	8.8 $\pm$ 1.4	7.7 $\pm$ 1.1	8.1 $\pm$ 1.3	0.403	0.75
S	11.7 $\pm$ 1.2	10.6 $\pm$ 1.2	11.6 $\pm$ 1.4	9.8 $\pm$ 1.5	3.035	0.034

**Fig. 4** Group-level average waveforms and scalp topographies of N1, P2, and P3 at four conditions (A, Av, As, and Avs) within auditory target group. Group-level average waveforms at A, Av, As, and Avs conditions are presented using red, green, blue, and black lines. The gray region in each panel shows the time interval to measure ERP peaks. *Left panel:* Group-level average waveforms

(measured at Fz) and scalp topographies of N1 at four conditions. *Middle panel:* Group-level average waveforms (measured at Cz) and scalp topographies of P2 at four conditions. *Right panel:* Group-level average waveforms (measured at Pz) and scalp topographies of P3 at four conditions

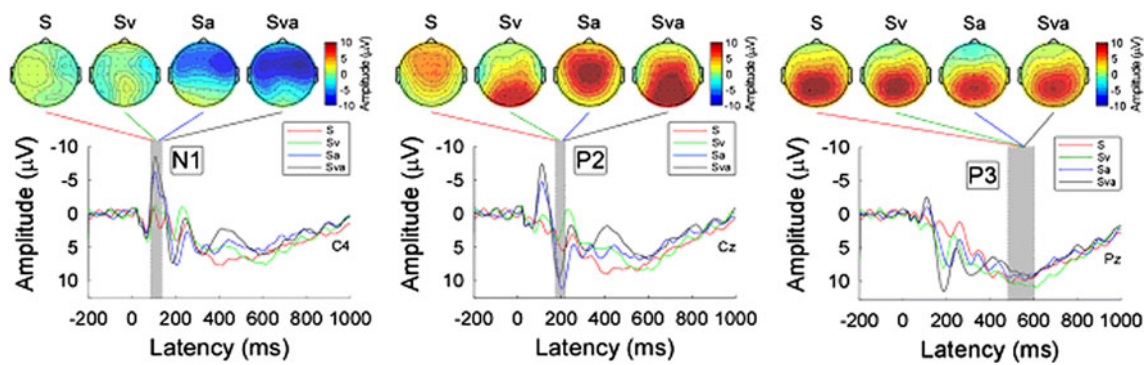


Fig. 5 Group-level average waveforms and scalp topographies of N1, P2, and P3 at four conditions (S, Sv, Sa, and Sva) within somatosensory target group. Group-level average waveforms at S, Sv, Sa, and Sva conditions are presented using red, green, blue, and black lines. The gray region in each panel shows the time interval to measure ERP peaks. *Left panel:* Group-level average waveforms

(measured at C4) and scalp topographies of N1 at four conditions. *Middle panel:* Group-level average waveforms (measured at Cz) and scalp topographies of P2 at four conditions. *Right panel:* Group-level average waveforms (measured at Pz) and scalp topographies of P3 at four conditions

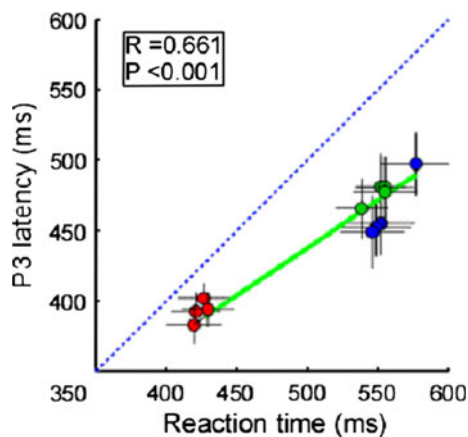


Fig. 6 Correlations between behavioral data (reaction time) and ERP data (P3 latency). Each red, green, and blue point represents the values from one condition within visual, auditory, and somatosensory target groups respectively. Vertical and horizontal error bars represent, for each condition, the variance across subjects (expressed as SEM). Green solid lines represent the best linear fit. Significant correlation was observed when examining the relationship between RTs and P3 latencies ($R = 0.661$, $P < 0.001$). Blue dashed line represents the identity line. Note that the RTs are significantly longer than P3 latencies ($P < 0.001$, two-tailed t test)

found in the bimodal and trimodal trials in the auditory target condition, whereas faster responses were obtained when tactile targets were presented in a bimodal and trimodal context. Both the auditory and tactile target groups revealed a significant difference in unimodal vs. bimodal and bimodal vs. trimodal comparisons.

Perceptual load and top-down modulation

Based on the findings of this study, it is highly likely that the perceptual load across different sensory inputs could be remarkably different, to the point of causing a significant

difference in the computational efficiency during working memory updating. In this study, we found that reaction time was significantly correlated with the P3 latency ($R = 0.661$, $P < 0.001$) (Fig. 6). Therefore, we speculated that efficiency in the target detection task was affected by top-down modulation.

The ERP results from the current experiment indicated that most of the ERP components were differentiated among unimodal, bimodal and trimodal target stimuli within each target modality group, especially with respect to the amplitude of N1, P2 and P3. The amplitude of N1 was significantly different among the three target modality groups and within each target modality group. In the visual target group, the N1 amplitudes in the multimodal conditions were generally larger than in the unimodal conditions ($V = -3.4 \mu V$, $Va = -5.4 \mu V$, $Vs = -3.6 \mu V$, $Vas = -5.5 \mu V$; $F = 3.361$, $P = 0.023$). Similar to the visual target group, in both the auditory and somatosensory target groups, the multimodal conditions had larger N1 amplitudes than the unimodal condition ($A = -10.7 \mu V$, $As = -10.9 \mu V$, $Av = -11.3 \mu V$, $Asv = -11.7 \mu V$; $F = 2.952$, $P = 0.038$) ($S = -2 \mu V$, $Sa = -8 \mu V$, $Sv = -3.8 \mu V$, $Sva = -10 \mu V$; $F = 17.33$, $P < 0.001$) (Table 1). It was known that the amplitude of N1 can be influenced by selective attention (Luck et al. 2005). The larger N1 amplitude indicated that more attention during target detection was allocated (Luck et al. 2000) to multimodal stimuli. In this experiment, the differences in the amplitude of N1 may reflect differences in attention levels among all the target conditions.

The amplitude of P2 was significantly different among the three target modalities and within the visual and somatosensory target modality groups. In the visual target group, the multimodal conditions generally had larger P2 amplitude than the unimodal conditions ($V = 3.61 \mu V$,

$V_a = 6.6 \mu\text{V}$, $V_s = 4.48 \mu\text{V}$, $V_{as} = 7.3 \mu\text{V}$; $F = 3.567$, $P = 0.018$). However, for the auditory target group, the difference between the multimodal and unimodal conditions was not significant ($A = 5.1 \mu\text{V}$, $A_s = 6.9 \mu\text{V}$, $A_v = 4.6 \mu\text{V}$, $A_{sv} = 4.8 \mu\text{V}$; $F = 0.934$, $P = 0.429$). The somatosensory target group had significantly larger amplitudes in the multimodal conditions than in the unimodal conditions ($S = 7.1 \mu\text{V}$, $S_a = 12.6 \mu\text{V}$, $S_v = 6.9 \mu\text{V}$, $S_{va} = 11.6 \mu\text{V}$; $F = 5.607$, $P = 0.002$) (Table 1).

The amplitude of P2 was found to be associated with a cognitive matching system, which compared sensory inputs with stored memory. The bottom-up sensory-specific information can be combined with top-down task relative information to accelerate or decelerate the detection processing within each target modality group (Treisman and Sato 1990; Luck and Hillyard 1994).

The amplitude of P3 was significantly different among the three target modality groups and within the somatosensory target modality groups. In the visual target group, the amplitude of P3 was not significantly different in the multimodal conditions than in the unimodal condition ($V = 13.3 \mu\text{V}$, $V_a = 9.9 \mu\text{V}$, $V_s = 12.4 \mu\text{V}$, $V_{as} = 10.3 \mu\text{V}$; $F = 0.93$, $P = 0.43$). In the auditory target group, the amplitude of P3 was not significantly different in the multimodal conditions than in the unimodal condition ($A = 9.4 \mu\text{V}$, $A_s = 8.8 \mu\text{V}$, $A_v = 7.7 \mu\text{V}$, $A_{sv} = 8.1 \mu\text{V}$; $F = 0.4$, $P = 0.75$). However, the amplitude of P3 in the somatosensory target group was significantly smaller in the multimodal than the unimodal target condition ($S = 11.7 \mu\text{V}$, $S_a = 10.6 \mu\text{V}$, $S_v = 11.6 \mu\text{V}$, $S_{va} = 9.8 \mu\text{V}$; $F = 3.035$, $P = 0.034$) (Table 1).

The amplitude of P3, which represented the allocation of attention during the detection of target stimuli (Luck et al. 2000; Rugg et al. 1987), may represent a facilitation effect when comparing the multimodal with the unimodal target stimuli within each target group. However, the amplitude of P3, which was influenced by the uncertainty of target detection and allocation of cognitive resources (Isreal et al. 1980; Johnson 1984), could indicate an inhibition effect when comparing the multimodal with unimodal target stimuli within each target group.

Dynamic characteristics of multisensory facilitation and inhibition

The neural representations of sensory signals encode information from the external world in a hierarchical fashion. The cognitive processing of multisensory signals can be differentiated in two key stages: (1) early automatic filtering of salient information and (2) late competitive selection of relevant neural representations (Desimone and Duncan 1995; Engel et al. 2001; Knudsen 2007; Miller and

Cohen 2001). These two stages, which are often referred to as the key concepts of bottom-up and top-down modulation respectively, have been validated by a series of previous studies on visual modality (Fallah et al. 2007; Magosso 2010; Reynolds and Chelazzi 2004), auditory modality (Bidet-Caulet et al. 2010; Han et al. 2007; Murray et al. 2005; van Atteveldt et al. 2007), and somatosensory modality (Hikita M et al. 2008; Mesulam 1998; Werkhoven et al. 2009). The coexistence of multisensory facilitation and inhibition, which occurs at these two stages respectively, plays a key role for understanding multisensory interaction.

Multisensory facilitation is thought to act at the level of the early automatic salience filtering stage, and functions to improve working memory performance (Stein and Meredith 1993). This facilitation process can be explained by Bayesian statistics principles. Bayesian statistics principles have shown that the imperfect estimate of occurrence obtained from a single sensory input can be improved by taking into account the probabilities of signals in other sensory channels. Therefore, the uncertainty of imperfect and noisy sensory input can be minimized by combining the probabilities of multiple sensory signals to refine sensory estimates (Kording and Wolpert 2006).

Multisensory inhibition is thought to act at the level of the late competitive selection stage, thus influencing working memory performance (Knudsen 2007). This late competitive selection of neural representation can be explained by the competition of stimulus signals in working memory, which is possibly controlled by a neural network containing a special type of inhibitory neuron, which can inhibit and compete with neighboring neurons (Knudsen 2007).

In this study, we observed that RTs for auditory stimuli were significantly shorter when presented in a unimodal condition than when accompanied by non-target stimuli belonging to another modality ($A < A_v$, $P = 0.025$; $A < A_s$, $P = 0.004$; $A < A_{sv}$, $P < 0.001$; two tailed paired t test) (Fig. 2). In addition, we found a significant correlation between RTs and P3 latency ($R = 0.64$, $P < 0.001$) within the auditory target group. Since P3 is an important component of ERPs that mainly reflects the process of working memory updating (Verleger 1991; Vogel et al. 1998; Li et al. 2011; Taylor 2007), we speculate that working memory updating played an important role in the difference of RTs, which were directly related to the difficulty of the task. Particularly, the difficulty for the participant in detecting the target event (A) was varied because of the presence of the multisensory background events (v , s , and vs) (Bresciani et al. 2008). As discussed above, early automatic salience filters for multimodal auditory target detection would mainly contribute to a facilitation effect compared with processing of unimodal

auditory targets. The results of amplitude of N1 and P2 in the current experiment suggests a weaker facilitation effect in the multimodal condition compared with the unimodal condition during the early feature identification stage in the auditory target group. Similarly, the late competitive selection of neural representation would mainly lead to an inhibition effect. This can be seen by combining the ERP results regarding the amplitude of P3 in the current experiment, suggesting a weak inhibition effect in the multimodal target condition compared with the unimodal condition during the late competitive selection of neural representation in the auditory target group. The combination of these two effects (facilitation and inhibition) would affect the attention-mediated top-down control (Bidet-Calet et al. 2010). For auditory targets, these results imply a weak multisensory inhibition and a weaker multisensory facilitation, thus demonstrating an overall multisensory inhibition. This overall multisensory inhibition would contribute to the difficulty of the task, making it harder for the participant to detect the target in the multimodal conditions than in the unimodal condition. This increase in difficulty would be represented by significantly longer RTs in the multimodal conditions.

For the somatosensory target condition, the RTs for unimodal stimuli were significantly longer than those for multimodal stimuli ($S > S_a$, $P = 0.029$; $S > S_v$, $P < 0.001$; $S > S_{va}$, $P < 0.001$; two tailed paired t test) (Fig. 2). As in the auditory target condition, significant correlations between RTs and P3 latency ($R = 0.62$, $P < 0.001$), were found within the somatosensory target group. Again, as in the auditory target condition, the difficulty for the participant in detecting the target event (S) would be varied because of the presence of multisensory background events as indicated by the amplitudes of N1, P2 and P3. However, there would be a strong multisensory inhibition and an even stronger multisensory facilitation thus resulting in an overall multisensory facilitation. This overall multisensory facilitation would contribute to improved perceptual processing and would make multimodal detection easier than unimodal detection (Lugo et al. 2008), as indicated by significantly shorter RTs in the multimodal conditions.

For the visual target condition, RTs were not significantly different between the unimodal target and multimodal target conditions (V vs. V_a , $P = 0.457$; V vs. V_s , $P = 0.464$; V vs. V_{as} , $P = 0.369$; two tailed paired t test) (Fig. 2). Interestingly, significant correlations were still observed between RTs and P3 latency ($R = 0.698$, $P < 0.001$) within the visual target group. Unlike both the auditory and somatosensory target conditions, difficulty in detecting the target event (V) did not significantly vary with the presence of multisensory background events (a, s, and as) (van Wassenhove et al. 2008). It is possible that the

combination of multisensory facilitation and inhibition cancelled one another out, thus resulting in a similar level of difficulty during working memory updating.

Altogether our results indicate that dynamic changes in target recognition are modulated by task difficulty (top-down load) during working memory updating, which represents an integrated effect of early facilitation and late inhibition phenomena. A weak multisensory inhibition and a weaker multisensory facilitation would contribute to an overall multisensory inhibition of auditory target detection because of the multisensory non-target background events. In contrast, a strong multisensory inhibition and a stronger multisensory facilitation would contribute to an overall

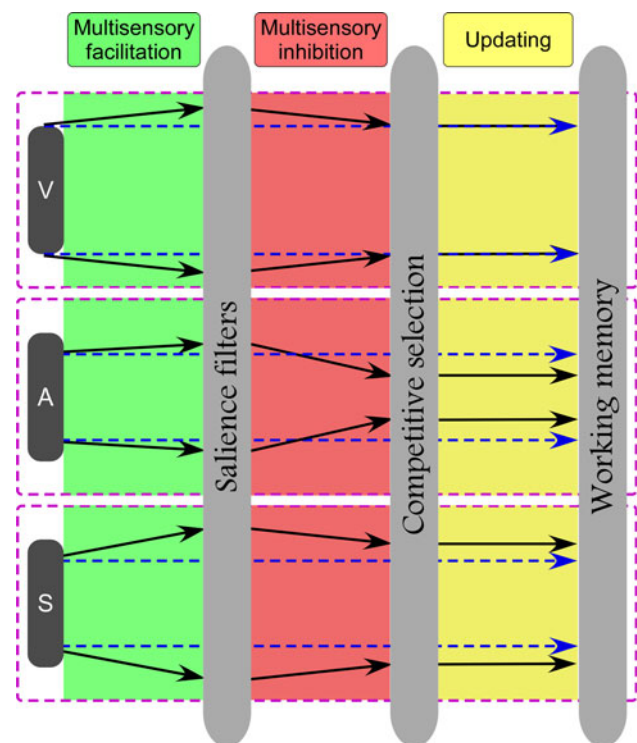


Fig. 7 Dynamic model for modality-specific cognitive processing. Across all three modalities (vision, audition, and somatosensory), this model would be comprised of three main parts: (1) Automatical salience filters; (2) Competitive selection; (3) Working memory updating. Blue dashed lines and black solid lines represent modality specific information flow without and with the influence of multisensory background events. Note that an increasing distance between two lines (either dashed or solid) for each modality represents a facilitation effect, and a decreasing distance between two lines for each modality represents an inhibition effect. *Top panel:* Similar multisensory facilitation and inhibition cancel out with each other, thus contributing to a balanced effect for visual target. *Middle panel:* The coexistence of appropriate multisensory facilitation and stronger inhibition, which occurred at early automatically salience filters and late competitive selection of neural representation, contributed to an overall multisensory inhibition for auditory target. *Bottom panel:* The coexistence of stronger multisensory facilitation and appropriate inhibition contributed to an overall multisensory facilitation for somatosensory target

multisensory facilitation of somatosensory target detection because of the multisensory background events. In addition, moderate levels of multisensory facilitation and inhibition were balanced in visual target detection among four visual target conditions.

Conclusions

While audition was affected by an overall multisensory inhibition, somatosensation was enhanced by an overall multisensory facilitation due to multisensory non-target background events. Based on previous findings regarding allocation of attention during multisensory processing (Knudsen 2007), we propose to extend the currently accepted model of dynamic processing of multisensory integration. This integration would determine the assembly of three main “modules”: (1) Automatic salience filters, (2) Competitive selection, and (3) Working memory updating (Fig. 7).

Considering the dynamic features of multimodal processing, the effect of both multisensory facilitation for somatosensory targets and multisensory inhibition for auditory targets may be caused by the interplay of multisensory facilitation and inhibition, which occur at both the entry level of automatic filtering of salient information, and during late competitive selection of neural representation respectively (Desimone and Duncan 1995; Engel et al. 2001; Knudsen 2007; Miller and Cohen 2001) (Fig. 7).

Acknowledgments The authors are grateful to Dr H. Z. Qi, and Ms. X. W. An for their help on experimental preparation and insightful comments on the study.

References

- Arabzadeh E, Clifford CW, Harris JA (2008) Vision merges with touch in a purely tactile discrimination. *Psychol Sci* 19(7): 635–641
- Beim Graben P, Gerth S, Vasissth S (2008) Towards dynamical system models of language-related brain potentials. *Cogn Neurodyn* 2(3):229–255
- Bidet-Caulet A, Mikyska C, Knight RT (2010) Load effects in auditory selective attention: evidence for distinct facilitation and inhibition mechanisms. *Neuroimage* 50(1):277–284
- Bresciani JP, Dammeier F, Ernst MO (2008) Tri-modal integration of visual, tactile and auditory signals for the perception of sequences of events. *Brain Res Bull* 75(6):753–760
- Chen YC, Yeh SL (2009) Catch the moment: multisensory enhancement of rapid visual events by sound. *Exp Brain Res* 198(2–3): 209–219
- Colavita FB (1974) Human sensory dominance. *Percept Psychophys* 16(2):409–412
- Desimone R, Duncan J (1995) Neural mechanisms of selective visual attention. *Annu Rev Neurosci* 18:193–222
- Driver J, Noesselt T (2008) Multisensory interplay reveals crossmodal influences on ‘sensory-specific’ brain regions, neural responses, and judgments. *Neuron* 57(1):11–23
- Driver J, Spence C (1998) Cross-modal links in spatial attention. *Philos Trans R Soc Lond B Biol Sci* 353(1373):1319–1331
- Engel AK, Fries P, Singer W (2001) Dynamic predictions: oscillations and synchrony in top-down processing. *Nat Rev Neurosci* 2(10):704–716
- Fallah M, Stoner GR, Reynolds JH (2007) Stimulus-specific competitive selection in macaque extrastriate visual area V4. *P Natl Acad Sci USA* 104(10):4165–4169
- Feng T, Qiu Y, Zhu Y, Tong S (2008) Attention rivalry under irrelevant audiovisual stimulation. *Neurosci Lett* 438(1):6–9
- Guntekin B, Basar E (2010) A new interpretation of P300 responses upon analysis of coherences. *Cogn Neurodyn* 4(2):107–118
- Han JH, Kushner SA, Yiu AP, Cole CJ, Matynia A, Brown RA, Neve RL, Guzowski JF, Silva AJ, Josselyn SA (2007) Neuronal competition and selection during memory formation. *Science* 316(5823):457–460
- Hartcher-O’Brien J, Gallace A, Krings B, Koppen C, Spence C (2008) When vision ‘extinguishes’ touch in neurologically-normal people: extending the Colavita visual dominance effect. *Exp Brain Res* 186(4):643–658
- Hecht D, Reiner M (2009) Sensory dominance in combinations of audio, visual and haptic stimuli. *Exp Brain Res* 193(2):307–314
- Hikita M, Fuke S, Ogino M, Minato T, Asada M (2008) Visual attention by saliency leads cross-modal body representation. *Int C Devel Learn* 1:157–162
- Isreal JB, Chesney GL, Wickens CD, Donchin E (1980) P300 and tracking difficulty: evidence for multiple resources in dual-task performance. *Psychophysiology* 17(3):259–273
- Jiang W, Jiang H, Stein BE (2002) Two corticotectal areas facilitate multisensory orientation behavior. *J Cogn Neurosci* 14(8):1240–1255
- Johnson R Jr (1984) P300: a model of the variables controlling its amplitude. *Ann NY Acad Sci* 425:223–229
- Knudsen EI (2007) Fundamental components of attention. *Annu Rev Neurosci* 30:57–78
- Koppen C, Spence C (2007) Seeing the light: exploring the Colavita visual dominance effect. *Exp Brain Res* 180(4):737–754
- Koppen C, Levitan CA, Spence C (2009) A signal detection study of the Colavita visual dominance effect. *Exp Brain Res* 196(3): 353–360
- Kording KP, Wolpert DM (2006) Bayesian decision theory in sensorimotor control. *Trends Cogn Sci* 10(7):319–326
- Li Y, Hu Y, Liu T, Wu D (2011) Dipole source analysis of auditory P300 response in depressive and anxiety disorders. *Cogn Neurodyn* 5(2):221–229
- Luck SJ (2005) An introduction to the event-related potential technique. MIT Press, Cambridge, MA
- Luck SJ, Hillyard SA (1994) Electrophysiological correlates of feature analysis during visual search. *Psychophysiology* 31(3):291–308
- Luck SJ, Woodman GF, Vogel EK (2000) Event-related potential studies of attention. *Trends Cogn Sci* 4(11):432–440
- Lugo JE, Doti R, Wittich W, Faubert J (2008) Multisensory integration: central processing modifies peripheral systems. *Psychol Sci* 19(10):989–997
- Magosso E (2010) Integrating information from vision and touch: a neural network modeling study. *Ieee T Inf Technol B* 14(3): 598–612
- McDonald JJ, Teder-Salejarvi WA, Di Russo F, Hillyard SA (2003) Neural substrates of perceptual enhancement by cross-modal spatial attention. *J Cogn Neurosci* 15(1):10–19
- Meredith MA, Stein BE (1986) Spatial factors determine the activity of multisensory neurons in cat superior colliculus. *Brain Res* 365(2):350–354

- Mesulam MM (1998) From sensation to cognition. *Brain* 121(Pt 6): 1013–1052
- Miller EK, Cohen JD (2001) An integrative theory of prefrontal cortex function. *Annu Rev Neurosci* 24:167–202
- Molholm S, Ritter W, Murray MM, Javitt DC, Schroeder CE, Foxe JJ (2002) Multisensory auditory-visual interactions during early sensory processing in humans: a high-density electrical mapping study. *Brain Res Cogn Brain Res* 14(1):115–128
- Murray MM, Molholm S, Michel CM, Heslenfeld DJ, Ritter W, Javitt DC, Schroeder CE, Foxe JJ (2005) Grabbing your ear: rapid auditory-somatosensory multisensory interactions in low-level sensory cortices are not constrained by stimulus alignment. *Cereb Cortex* 15(7):963–974
- Ocelli V, O'Brien JH, Spence C, Zampini M (2010) Assessing the audiotactile Colavita effect in near and rear space. *Exp Brain Res* 203(3):517–532
- Odgaard EC, Ariei Y, Marks LE (2004) Brighter noise: sensory enhancement of perceived loudness by concurrent visual stimulation. *Cogn Affect Behav Neurosci* 4(2):127–132
- Reynolds JH, Chelazzi L (2004) Attentional modulation of visual processing. *Annu Rev Neurosci* 27:611–647
- Rowland BA, Quessy S, Stanford TR, Stein BE (2007) Multisensory integration shortens physiological response latencies. *J Neurosci* 27(22):5879–5884
- Rugg MD, Milner AD, Lines CR, Phalp R (1987) Modulation of visual event-related potentials by spatial and non-spatial visual selective attention. *Neuropsychologia* 25(1A):85–96
- Schinkel S, Marwan N, Kurths J (2007) Order patterns recurrence plots in the analysis of ERP data. *Cogn Neurodyn* 1(4):317–325
- Sinnett S, Soto-Faraco S, Spence C (2008) The co-occurrence of multisensory competition and facilitation. *Acta Psychol (Amst)* 128(1):153–161
- Stein BE, Meredith MA (1993) *The merging of the senses*. MIT Press, Cambridge, MA
- Tanaka E, Inui K, Kida T, Miyazaki T, Takeshima Y, Kakigi R (2008) A transition from unimodal to multimodal activations in four sensory modalities in humans: an electrophysiological study. *BMC Neurosci* 9:116
- Taylor JG (2007) On the neurodynamics of the creation of consciousness. *Cogn Neurodyn* 1(2):97–118
- Teder-Salejari WA, Di Russo F, McDonald JJ, Hillyard SA (2005) Effects of spatial congruity on audio-visual multimodal integration. *J Cogn Neurosci* 17(9):1396–1409
- Treisman A, Sato S (1990) Conjunction search revisited. *J Exp Psychol Hum Percept Perform* 16(3):459–478
- van Atteveldt NM, Formisano E, Goebel R, Blomert L (2007) Top-down task effects overrule automatic multisensory responses to letter-sound pairs in auditory association cortex. *Neuroimage* 36(4):1345–1360
- Van Damme S, Crombez G, Spence C (2009) Is visual dominance modulated by the threat value of visual and auditory stimuli? *Exp Brain Res* 193(2):197–204
- van Wassenhove V, Buonomano DV, Shimojo S, Shams L (2008) Distortions of subjective time perception within and across senses. *Plos One* 3(1):e1437
- Verleger R (1991) Event-related potentials and cognition—a critique of the context updating hypothesis and an alternative interpretation of P3. *Behav Brain Sci* 14(4):732
- Vogel EK, Luck SJ, Shapiro KL (1998) Electrophysiological evidence for a postperceptual locus of suppression during the attentional blink. *J Exp Psychol Hum Percept Perform* 24(6): 1656–1674
- Werkhoven PJ, van Erp JBR, Philippo TG (2009) Counting visual and tactile events: the effect of attention on multisensory integration. *Atten Percept Psycho* 71(8):1854–1861