Biological Psychology 149 (2020) 107805

Contents lists available at ScienceDirect

Biological Psychology

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

# Multiple phases of cross-sensory interactions associated with the audiovisual bounce-inducing effect

Song Zhao, Yajie Wang, Chengzhi Feng*, Wenfeng Feng*

*Department of Psychology, School of Education, SooChow University, Suzhou, Jiangsu, 215123, China*

ARTICLE INFO

*Keywords:*
Bistable perception
streaming/bouncing
Audiovisual bounce-inducing effect
Event-related potential
Cross-modal interaction

ABSTRACT

Using event-related potential (ERP) recordings, the present study investigated the cross-modal neural activities underlying the audiovisual bounce-inducing effect (ABE) via a novel experimental design wherein the audiovisual bouncing trials were induced solely by the ABE. The within-subject (percept-based) analysis showed that early cross-modal interactions within 100–200 ms after sound onset over fronto-central and occipital regions were associated with the occurrence of the ABE, but the cross-modal interaction at a later latency (ND250, 220–280 ms) over fronto-central region did not differ between ABE trials and non-ABE trials. The between-subject analysis indicated that the cross-modal interaction revealed by ND250 was larger for subjects who perceived the ABE more frequently. These findings suggest that the ABE is generated as a consequence of the rapid interplay between the variations of early cross-modal interactions and the general multisensory binding predisposition at an individual level.

## 1. Introduction

As events occurring in the natural environment usually bring us with signals in more than one sensory modality, our brain needs to integrate these different sensory signals appropriately in order to generate meaningful percepts and then adaptive behaviors. Moreover, multisensory integration sometimes occurs in a striking way such that the perceptual outcome of signals from one sensory modality could be extremely influenced by information from another modality (for review, see Driver & Noesselt, 2008). Within the audiovisual domain, one of the most compelling examples is the effect of sound on the streaming/bouncing visual motion perception introduced by Sekuler, Sekuler, and Lau (1997). Specifically, if two identical visual disks move toward one another on a two-dimensional display, they are more likely to be perceived as “streaming through” than “bouncing off” each other after their coincidence, although the bouncing percept still occurs occasionally (Bertenthal, Banton, & Bradbury, 1993; Metzger, 1934; Sekuler & Sekuler, 1999; Watanabe & Shimojo, 1998; Zhao et al., 2017). When presenting a transient auditory stimulus at the coincident moment of the two disks, however, the incidence of bouncing percept will be dramatically increased (Sekuler et al., 1997). As this effect of audition on visual motion perception was often referred to as “audiovisual bounce-inducing effect (ABE)” in recent studies (e.g., Grassi & Casco, 2009, 2010, 2012; Maniglia, Grassi, Casco, & Campana, 2012; Zhao, Wang, Xu, Feng, & Feng, 2018), the current study termed this phenomenon “ABE” for convenience.

The ABE phenomenon has been consistently observed in numerous behavioral studies (Watanabe & Shimojo, 2001a, 2001b; Shimojo & Shams, 2001; Remijn, Ito, & Nakajima, 2004; Kawabe & Miura, 2006; Dufour, Touzalin, Moessinger, Brochard, & Després, 2008; Grassi & Casco, 2009, 2010, 2012; Grove & Sakurai, 2009; Grove, Ashton, Kawachi, & Sakurai, 2012; Grove, Robertson, & Harris, 2016; Roudaia, Sekuler, Bennett, & Sekuler, 2013; Zeljko & Grove, 2016; Parise & Ernst, 2017) and has been widely utilized as a case of cross-modal interaction for investigating many other scientific issues, such as recalibration of audiovisual simultaneity (Fujisaki, Shimojo, Kashino, & Nishida, 2004), oscillatory synchronization in cortical networks (Hipp, Engel, & Siegel, 2011), effect of mental imagery on multisensory perception (Berger & Ehrsson, 2013, 2017), and attentional modulation on temporal binding of audiovisual stimuli (Donohue, Green, & Woldorff, 2015). However, the neural mechanisms responsible for the ABE are only beginning to be understood. A functional magnetic resonance imaging (fMRI) study conducted by Bushara et al. (2003) was the first to explore the neural correlates of the ABE. By comparing brain activities between trials on which bouncing percept was reported (i.e. bouncing trials) and streaming trials in the audiovisual motion display, they found enhanced hemodynamic response in a series of high-level multisensory regions (e.g. prefrontal and posterior parietal cortices) but decreased activation

* Corresponding author.
*E-mail addresses:* psyfrank@163.com (C. Feng), fengwfly@gmail.com (W. Feng).

https://doi.org/10.1016/j.biopsycho.2019.107805
Received 7 March 2019; Received in revised form 15 October 2019; Accepted 28 October 2019
Available online 02 November 2019

in unisensory auditory and visual cortices on the audiovisual bouncing than streaming trials, implying the competition between multisensory and unisensory brain regions might contribute to the occurrence of the ABE (Bushara et al., 2003). A similar pattern of results was reported in a subsequent event-related magnetoencephalograph (MEG) study using the same percept-based comparison method (Zvyagintsev, Nikolaev, Sachs, & Mathiak, 2011). Besides, an electroencephalograph (EEG) study using the identical percept-based comparison method but focusing on EEG oscillatory synchronization found that beta-band synchronization across frontal, parietal, and occipital cortices, and gamma-band synchronization across central and temporal regions were higher on the audiovisual bouncing than streaming trials, indicating the oscillatory coherence across large-scale brain networks are also involved in the triggering of the ABE (Hipp et al., 2011). More recently, an event-related potentials (ERP) study using the method of isolating cross-modal neural activities showed that two early cross-modal ERP components, the fronto-central positivity (PD170, 125–175 ms after sound onset) and the occipital positivity (PD190, 180–200 ms), were significantly larger on the audiovisual bouncing than streaming trials. Indeed, the earliest PD170 component was completely absent on the audiovisual streaming trials wherein the ABE obviously did not occur (Zhao et al., 2018). In contrast, the later cross-modal negativity (ND250, 220–280 ms) was smaller on the audiovisual bouncing than streaming trials. Given the early latency of the differences in the cross-modal neural activities observed on a percept-related basis, it was concluded that cross-modal interactions at perceptual stage of processing underlie the occurrence of ABE phenomenon (Zhao et al., 2018).

As reviewed above, previous neuroscience efforts mainly adopted the approach of comparing the audiovisual bouncing with streaming trials to investigate the neural substrates of the ABE [but see Maniglia et al. (2012) who utilized transcranial magnetic stimulation (TMS) to interrupt the function of posterior parietal cortex and compared the ABE magnitude between conditions with and without TMS]. This percept-based analysis assumed that bouncing trials in audiovisual streaming/bouncing display represented the trials on which the transient sound influenced the visual motion perception (i.e. the ABE occurred), whereas the audiovisual streaming trials reflected the trials that the sound failed to bias the visual perception (i.e. the ABE did not occur; Bushara et al., 2003; Zvyagintsev et al., 2011; Hipp et al., 2011; Zhao et al., 2018). However, it is noteworthy that audiovisual bouncing trials in these studies might include not only the trials on which the ABE occurred, but might also include some trials on which bouncing percept would be reported even if without the influence of the brief sound. This is because the streaming/bouncing displays without sounds designed in these studies were subjectively bistable [i.e. ambiguous, except for the fMRI study conducted by Bushara et al. (2003)], which was characterized by certain trials (about 30% or more) being also perceived as bouncing event in visual-only display (e.g. Hipp et al., 2011; Zhao et al., 2018). Therefore, in order to investigate the exact time course of the neural mechanisms of the ABE that is unconfounded by the above-mentioned factor, it is necessary to further extract the audiovisual bouncing trials that were induced by the ABE only. Given this background, the present study used a novel experimental design in which almost all visual-only displays were perceived as streaming, thus the bouncing responses in audiovisual trials could be considered as resulting solely from the ABE. Based on this modification in paradigm, the present study used ERP recordings in conjunction with the method of isolating brain activities associated with cross-modal interactions, as well as the classic percept-based analysis to further investigate early cross-modal neural activities associated with the occurrence of the ABE.

It should also be noted that the neural substrates underlying the inter-individual variability in the tendency to perceive the ABE (i.e. the magnitude of ABE, characterized as the difference in the percentage of bouncing percept between audiovisual and visual-only displays) remain to be determined. Only one EEG study (Hipp et al., 2011), to our knowledge, had explored this issue and found that individuals with smaller difference in gamma-band synchronization across central and temporal brain areas between audiovisual bouncing and streaming trials tended to perceive the ABE more frequently. However, given that the ABE has been widely considered as an audiovisual cross-modal phenomenon (Berger & Ehrsson, 2013; Bushara et al., 2003; Dufour et al., 2008; Fujisaki et al., 2004; Remijn et al., 2004; Sanabria, Correa, Lupiáñez, & Spence, 2004; Scheier, Lewkowicz, & Shimojo, 2003; Zhou, Wong, & Sekuler, 2007, 2017; Donohue et al., 2015; Kawachi, 2016; Meyerhoff & Scholl, 2018; Meyerhoff, Merz, & Frings, 2018; Gobara, Yoshimura, & Yamada, 2018), it is currently unclear whether neural activities directly associated with cross-modal interaction contribute to the individual difference in behavioral ABE magnitude. Furthermore, in order to further understand the neural mechanisms of the ABE, it is also necessary to take into account the inter-individual variability in predisposition to perceive the ABE, which did not draw much attention from most previous neuroscience works (i.e. Bushara et al., 2003; Zvyagintsev et al., 2011; Maniglia et al., 2012; Zhao et al., 2018). Therefore, the second aim of the present study was to examine this question by comparing cross-modal neural activities as characterized by ERPs between subjects who were disposed to experience the ABE more frequently and those who perceived the ABE less frequently.

## 2. Methods

### 2.1. Participants

A total of 44 healthy paid subjects (29 females, mean age of 21 years, range 18–28 years) participated in the study after giving written informed consent as approved by the Human Research Protections Program of SooChow University. Each subject had normal or corrected-to-normal visual acuity as well as normal audition. They were all naive as to the detailed hypothesis of the experiment. All experimental procedures were in agreement with the Declaration of Helsinki.

### 2.2. Stimuli and task

The experiment was performed in a dimly lit and sound-attenuated chamber. Stimulus presentation was scripted using “Presentation” software (version 18.0, NeuroBehavioral Systems, Inc.). Visual motion stimuli were presented on a 27-inch LCD screen (ASUS PG279Q, resolution 1920 × 1080, refresh rate 120 Hz) on which the background color was set to gray (RGB: 127, 127, 127), and auditory stimuli were delivered by a pair of loudspeakers (HiVi X3) positioned at the left and right sides of the screen so that a single sound presented by the two speakers simultaneously would be perceived as coming from the center of the screen (Bertelson & Aschersleben, 1998). Participants sat in front of the screen with a viewing distance of approximately 85 cm, and were required to maintain their eyes fixated on a red cross (RGB: 255, 0, 0; 0.3° × 0.3° of visual angle), which was displayed at the center of the screen throughout each experimental block.

There were five stimulus conditions in the present experiment, which were labeled as “V”, “VA”, “A”, “N” and “Catch” conditions respectively for simplicity (see Fig. 1A). Firstly, the **V condition** in the present study referred to a modified version of the streaming/bouncing visual motion display where the velocity of the two moving disks was quite slow. Specifically, on the first frame, two identical black disks (RGB: 0, 0, 0; each 1.05° in diameter) were presented at the left and right sides of the screen, separated by 4.2° horizontally and placed both 3.46° above the central fixation for 50 ms. From frame 2 to frame 8, the two disks moved towards one another horizontally with uniform rectilinear motion. Each frame in the experiment appeared instantly after the disappearance of the preceding frame, and the duration of each frame was 50 ms [i.e. frame to frame stimulus onset asynchrony (SOA) was 50 ms]. From frame 9 to frame 11 (see Fig. 1B, frames highlighted in red), the two disks started to *partially* overlap with each other, and then became *completely* overlap (i.e. coincidence) at the onset moment

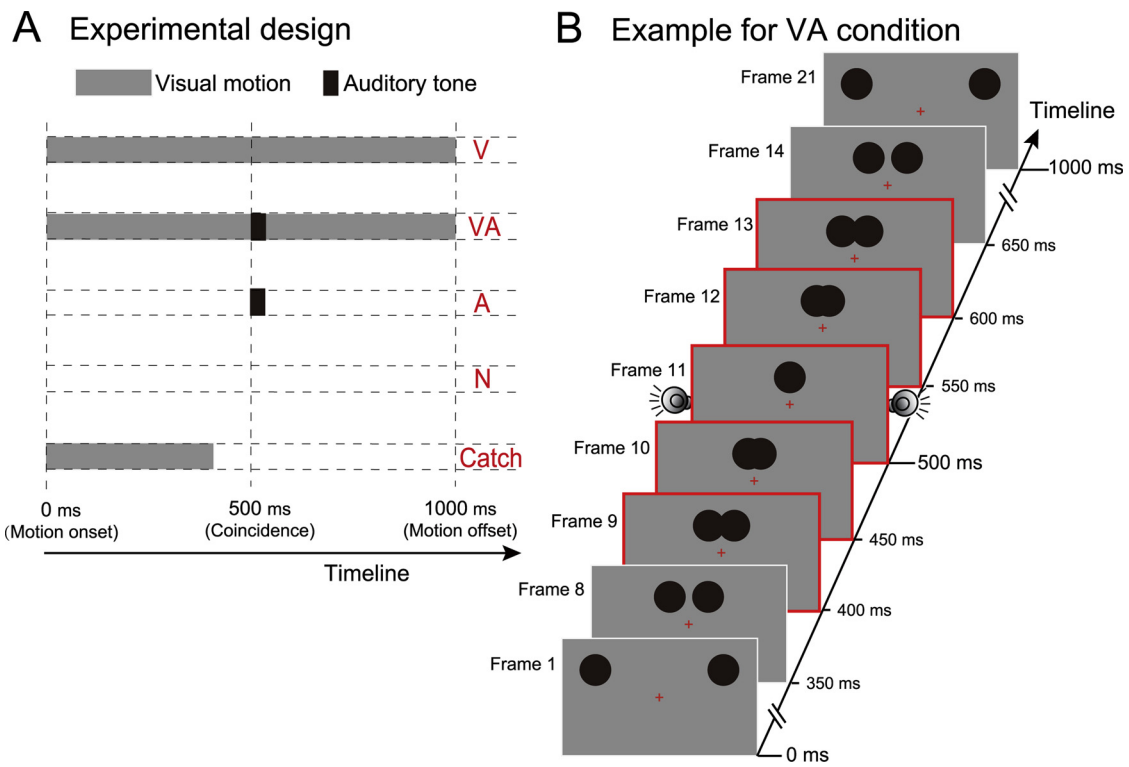


**Fig. 1.** (**A**) Overview of the five stimulus conditions designed for the experiment, which were labeled in red as V, VA, A, N, and Catch, respectively (see Stimuli and task section for details). The horizontal axis below (timeline) indicates the onset moments and durations of the visual motion sequence (gray bar) as well as the auditory tone (black bar). (**B**) Schematic illustration of the stimulus sequence exemplifying the VA condition, wherein two identical visual disks moved towards each other (frame 1–8), *partially* overlapped (frame 9–10), *completely* overlapped accompanied by the presentation of the auditory tone (frame 11), moved apart (frame 12–20) and then stopped at each other's starting points (frame 21), after which they disappeared. The visual motion took 1000 ms with a constant and slow velocity of 4.2°/s, and each frame presented instantly after the disappearance of the preceding frame. The solid axis on the right (timeline) denotes the onset moments of the key frames shown here. The frames highlighted in red here denote the frames on which the two visual disks overlapped partially or completely with each other, and were not highlighted when conducting the experiment.

of frame 11 (i.e. 500 ms after the frame 1 onset). From frame 12 through 21, the two disks gradually moved away from one another, then stopped at each other's starting point and finally disappeared. Given the initial 4.2° separation of the two disks, the 50 ms duration of each frame and a total of 21 frames as well, the visual motion took 1000 ms with a constant velocity of 4.2°/s (i.e. 0.21° per frame). Such a slow velocity was chosen because the results from several previous ABE studies suggested that a slow speed of the two moving disks could substantially reduce the ambiguity of the visual motion display, thus could make almost all visual-only trials be perceived as streaming event (c.f. Watanabe & Shimojo, 2001b, their Exp. 2 & 3; Bushara et al., 2003, their Exp. 1; Grassi & Casco, 2010, their Exp. 2; Grassi & Casco, 2012; Maniglia et al., 2012; Kawachi, 2016, his Exp. 1). The slow movement of two disks seems to make the detailed motion sequence near the coincidence of the two disks [i.e. from partial overlap to complete overlap and then to partial overlap again] be perceived more clearly, which might be a strong visual cue biasing towards the streaming percept in the visual-only display.

Secondly, the **VA condition** was comprised of the same visual motion display as that in V condition, as well as an auditory pure tone (800 Hz, approximately 70 dB at subjects' ears, 50 ms duration with 5 ms rise and fall ramps) delivered from the two loudspeakers simultaneously at the moment when the two visual disks overlapped completely (i.e. at the onset moment of frame 11, see Fig. 1B). Given that almost all trials in the V condition would be perceived as streaming, the occurrence of bouncing responses in the VA condition thus could be considered as originating solely from the ABE. Thirdly, in the **A condition**, no visual stimuli (except the fixation cross, the same below) were presented at all from frame 1 to frame 21, but the same auditory tone as that in VA condition was presented at the onset moment of frame 11 (Fig. 1A). Fourthly, the **N condition** presented neither visual nor auditory stimuli, but EEG signals in this condition were also recorded for further analysis. The reason for including this no-stimulus condition is detailed in the percept-based ERP analysis section (see below). Finally, in the **Catch condition**, the visual motion presented from frame 1 to frame 8 was identical to that in V condition, but no stimulus was presented from frame 9 to frame 21 (Fig. 1A). That is, the two visual disks moved towards one another then suddenly vanished just before their partial overlap, which would induce neither streaming nor bouncing percept. These catch trials were included in order to ensure that subjects responded veridically based on their perceptual outcome after the coincidence event (i.e. the complete overlap of two disks) happened instead of simply relying on guesswork before that event occurred (Zhao et al., 2017, 2018).

The five stimulus conditions occurred with equal probability in a randomized sequence in each block, and the inter-trial interval (ITI) varied from 1200 to 1600 ms randomly. The task for participants was to indicate whether the two visual disks appeared to "stream through" or "bounce off" one another after their coincidence (i.e. to report the perceptual outcome in V and VA conditions) by pressing one of two buttons ("F" and "J") on a keyboard. The response buttons for "streaming" and "bouncing" percepts were counterbalanced between subjects and no responses were required to the other three stimulus conditions where no post-coincidence visual motion was presented. The instructions emphasized response accuracy (i.e. responding veridically according to the subjective perceptual outcome) more than speed. The whole experiment consisted of 30 blocks of 60 trials each, so that a large number of 360 trials were recorded overall for each stimulus

condition. Participants were instructed to have a rest between blocks in order to relieve eye fatigue.

### 2.3. Electrophysiological recording and processing

The EEG was recorded continuously when subjects performed the behavioral task using the NeuroScan (SynAmps) acquisition system with "Scan" software (version 4.3, NeuroScan, Inc.). EEG signals were collected from 64 electrode sites based on an extended 10–20 system montage (for details, see Zhao et al., 2017). The horizontal electro-oculogram (HEOG) triggered by horizontal eye movements were recorded bipolarly via two electrodes placed at the outer canthi of the eyes, and vertical EOG induced by vertical eye movements and eye blinks were monitored bipolarly via two electrodes positioned above and below the left eye. The importance of fixation was emphasized to participants. The impedances of all electrodes were kept below 5 kΩ before EEG acquisition. The online EEG and EOG signals were amplified with a gain of 10,000, a band-pass filter of 0.05–100 Hz, and were continuously digitized with a sampling rate of 1000 Hz. All scalp electrode sites were referenced to the left mastoid electrode during data recording but were re-referenced offline to the algebraic average of the left and right mastoid electrodes.

In offline processing, the continuous EEG signals were firstly low-pass filtered digitally (30 Hz, 24 dB/octave) using a zero phase-shift FIR (finite impulse response) filter to remove high-frequency noise triggered by muscle activity or external electrical sources. EEG signals in V, VA, A and N conditions were then divided into 500-ms epochs time-locked to the onset of frame 11 (i.e. the coincidence moment of the two disks, also the onset moment of the auditory tone for VA and A conditions, Fig. 1 A & B) with a 100-ms pre-coincidence baseline. This segment of epochs was chosen because logically, audiovisual cross-modal interactions could occur only after both the visual and auditory stimuli have been presented (after the onset moment of the coincident sound). Automatic artifact rejection was performed according to a threshold of ± 60 μV for both EEG and EOG channels, in order to eliminate epochs contaminated by eye movements, eye blinks, muscle activity and amplifier blocking, leaving on average 331 ± 4 ($M \pm SE$) valid epochs in V condition, 329 ± 4 epochs in VA condition, 295 ± 7 in A condition and 303 ± 6 in. N condition, respectively. The remaining artifact-free epochs in each stimulus condition were baseline-corrected and then averaged separately to obtain corresponding ERP waveforms. ERP processing was carried out using "Scan" software (version 4.5).

### 2.4. Percept-based ERP analysis

ERP components associated with cross-modal interaction were isolated by calculating cross-modal difference (CMdiff) waveforms, which were obtained by subtracting the summed ERPs elicited by the unimodal V and A stimuli from ERPs evoked by the bimodal VA stimuli (c.f. Giard & Peronnet, 1999; Molholm et al., 2002; Fort, Delpuech, Pernier, & Giard, 2002; Teder-Sälejärvi, McDonald, Di Russo, & Hillyard, 2002; Teder-Sälejärvi, Di Russo, McDonald, & Hillyard, 2005; Talsma & Woldorff, 2005; Gondan & Röder, 2006; Bonath et al., 2007; Talsma, Doty, & Woldorff, 2007; Mishra, Martinez, Sejnowski, & Hillyard, 2007; Mishra, Martinez, & Hillyard, 2008, 2010; Li, Wu, & Touge, 2010; Senkowski, Saint-Amour, Höfle, & Foxe, 2011; Van der Burg, Talsma, Olivers, Hickey, & Theeuwes, 2011; Yang et al., 2013; Gao et al., 2014; Zhao et al., 2018). In order to examine whether the variations of early cross-modal neural activities are responsible for the occurrence of ABE, these cross-modal difference waveforms were calculated separately for *VA_bouncing trials* [**CMdiff_*bou*** = **VA_*bou* - (V + A–N)**] and *VA_streaming trials* [**CMdiff_*str*** = **VA_*str* - (V + A–N)**].

The N (no-stimulus) condition was included in these calculations in order to cancel out any pre-stimulus anticipatory ERP (such as the CNV, Walter, Cooper, Aldridge, McCallum, & Winter, 1964) that might extend into the post-stimulus period in all conditions (c.f. Talsma & Woldorff, 2005; Bonath et al., 2007; Mishra et al., 2007, 2008, 2010; Zhao et al., 2018). If the N condition were not included, such anticipatory ERPs would be added once but subtracted twice in these calculations, which might introduce a very early deflection (earlier than 100 ms after stimulus onset) that could be mistaken for a genuine cross-modal interaction (for details, see Teder-Sälejärvi et al., 2002, 2005; Talsma & Woldorff, 2005; Gondan & Röder, 2006). Moreover, the N, A (auditory-only) and V (visual-only) trials used for calculating cross-modal interactions on VA_bouncing trials (CMdiff_bou) and VA_streaming trials (CMdiff_str) were identical (c.f. Bonath et al., 2007; Mishra et al., 2007, 2008, 2010), because A and N conditions were task-irrelevant (no responses were required) and V condition was actually unambiguous (i.e. almost all V trials were perceived as streaming event, see Behavioral results section for details). In addition, only streaming trials in V condition were used to obtain ERPs in V condition. Accordingly, the differences between CMdiff_bou and CMdiff_str waveforms were algebraically equal to the differences when directly comparing VA_bouncing with VA_streaming trials. However, if ERP waveform on VA_bouncing trials were compared with that on VA_streaming trials directly, we could not have known whether the observed differences arise from cross-modal interactions and whether cross-modal interactions occur at all.

After calculating, the time windows and electrodes for measuring ERP components in cross-modal difference waveforms were selected ***a priori*** on the basis of the recent study conducted by Zhao et al. (2018) wherein similar behavioral task and data analysis procedure were performed. Specifically, the fronto-central positive difference component labeled as **PD170** was measured as mean amplitude within a time window of 125–175 ms after sound onset over a cluster of 10 fronto-central electrodes (FC3, FC1, FCz, FC2, FC4; C3, C1, Cz, C2, C4); The subsequent occipital positivity labeled as **PD190** was quantified as mean voltage during 180–200 ms over a cluster of 12 bilateral occipital sites (P7, P5, PO7, PO5, PO3, O1; P6, P8, PO4, PO6, PO8, O2); The fronto-central negativity labeled as **ND250** was measured as mean amplitude within 220–280 ms over the same 10 fronto-central electrode sites as PD170. ERPs occurring 300 ms after sound onset were not analyzed because neural activities related to cross-modal interaction occurring 300 ms after stimuli onset might be confounded with brain activities associated with decision-making and response preparation when calculating the cross-modal difference waveform (c.f. Gondan & Röder, 2006; Mishra et al., 2007, 2008, 2010; Cappe, Thut, Romei, & Murray, 2010; Cappe, Thelen, Romei, Thut, & Murray, 2012; Zhao et al., 2018).

To examine whether cross-modal interactions indexed by these ERP components were present or absent (i.e. whether the amplitudes of these cross-modal ERPs are statistically significant) on both VA_bouncing and VA_streaming trials, the mean amplitude of the bimodal ERP waveform was compared with that of the summed unimodal ERP waveform within each PD or ND interval separately for VA_bouncing and VA_streaming trials (c.f. Molholm et al., 2002; Talsma & Woldorff, 2005; Talsma et al., 2007). These comparisons were performed because each PD or ND component was actually the *difference* between the bimodal ERP waveform and the summed unimodal ERP waveform. Specifically, for VA_bouncing trials, the significance of each PD/ND was tested by repeated-measure ANOVA with a single factor of ERP type [VA_*bou* v.s. (V + A - N)]. For VA_streaming trials, similar one-way repeated-measure ANOVAs with the factor of ERP type [VA_*str* v.s. (V + A - N)] were performed. Note that the summed unimodal ERP waveform (V + A - N) used for comparisons are identical for VA_bouncing and VA_streaming trials. Moreover, the amplitudes of these PD/ND components that represented the magnitudes of cross-modal interactions were then compared directly between CMdiff_bou and CMdiff_str waveforms, in order to further investigate whether early cross-modal interactions underlie the occurrence of ABE. The mean amplitudes of each PD or ND component were thus subject to a

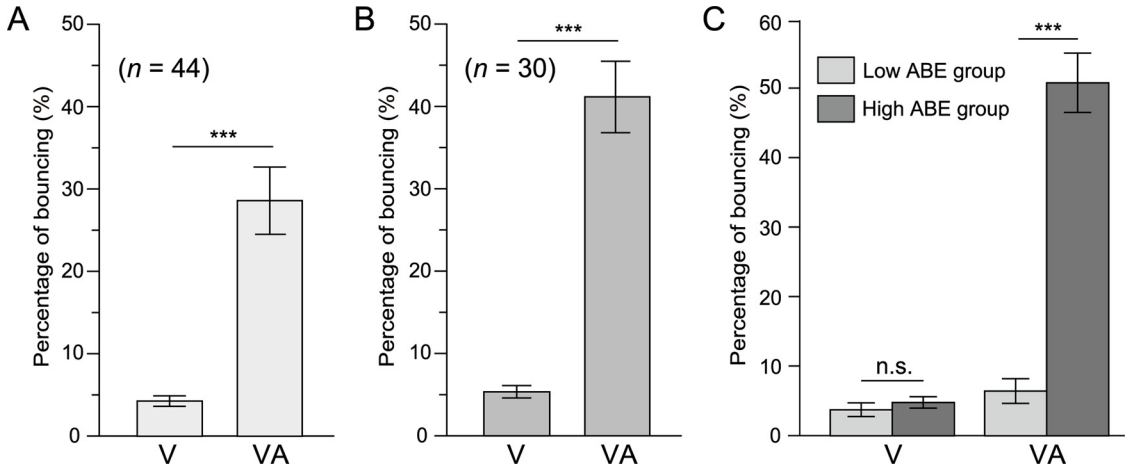


**Fig. 2.** Behavioral AEB effect. Mean percentages of bouncing percept in V (i.e. visual-only motion) and VA (i.e. audiovisual motion) conditions for all 44 participants (**A**) and for the 30 participants whose ERP data were analyzed on a percept-related basis (**B**, see percept-based ERP analysis section). The percentage of bouncing percept in V condition was extremely low, suggesting the bouncing responses in VA condition here could be approximately considered as deriving solely from the ABE. (**C**) Mean percentages of bouncing further depicted as functions of condition (V v.s. VA) and group (high v.s. low ABE group). Note that no group difference was found in V condition, indicating the two groups did not differ in perceptual or decision criteria when perceiving the unimodal visual motion. Error bars in all graphs represent ± 1 *SEM*; ***: $p < 0.0001$; n.s.: nonsignificant.

repeated-measure ANOVA with a single factor of perception (CMdiff_bou v.s. CMdiff_str), respectively. All *p*-values for ANOVA results were corrected using the Greenhouse-Geisser method. The electrode or hemisphere was not included as a factor in ANOVAs in the current study because we did not have enough prior knowledge about the lateralization of these cross-modal ERP components, and ANOVA with a factor of electrode or hemisphere in this case might increase the likelihood of type I error.

The above-mentioned percept-based analysis of cross-modal interactions was carried out for 30 subjects (19 females) whose bouncing and streaming percepts in VA condition were both more than a minimum of 30 trials, which was required to maintain an acceptable signal-to-noise ratio for ERP analysis (c.f. Capizzi, Correa, & Sanabria, 2013). Among these subjects, there were on average 135 ± 15 (*M* ± *SE*) VA_bouncing trials (41%) and 193 ± 15 VA_streaming trials (59%) left for obtaining respective ERP waveforms after artifact rejection, and the number of trials did not differ significantly from one another [$F(1, 29) = 3.95$, $p = 0.056$, $\eta_p^2 = 0.12$]. Data of the remaining 14 participants were excluded from this percept-based analysis due to extremely inadequate number (less than 30) of either VA_bouncing or VA_streaming trials, which might severely undermine the signal-to-noise ratio of ERP waveforms and, consequently, the outcomes of the percept-based analysis.

### 2.5. Between-subject ERP analysis

To investigate whether neural activities associated with cross-modal interactions are also responsible for the inter-individual variability in tendency to perceive the ABE, all forty-four participants were divided into two groups (22 in each) by a median split of the ABE magnitude, which was measured for each participant as the difference in percentage of bouncing percept between VA and V conditions (e.g. Sekuler et al., 1997; Watanabe & Shimojo, 2001a, 2001b; Grassi & Casco, 2009, 2010, 2012; Grove & Sakurai, 2009; Grove et al., 2012, 2016). Subjects in the *"high ABE" group* were disposed to perceive the ABE more frequently [i.e. had a higher ABE magnitude (*M* ± *SE*: 46.0 ± 4.3%)] whereas those in the *"low ABE" group* showed a very low or even no ABE magnitude (2.7 ± 1.0%). These two groups were equivalent in gender distribution and age (high ABE group: 7 males and 15 females, mean age of 21.8 years; low ABE group: 8 males and 14 females, mean age of 20.4 years). ERPs associated with cross-modal interactions were isolated for each subject by calculating the cross-modal difference waveform [**CMdiff = VA - (V + A–N)**] over all trials in each stimulus condition (i.e. without separation between VA_bouncing and VA_streaming trials) except the V condition under which only streaming trials were utilized. The time intervals and electrode sites selected for measuring ERP components (i.e. PD170, PD190 and ND250) in the CMdiff waveform were identical to those used for percept-based analysis (see above).

For statistical analysis, firstly, the mean amplitude of the bimodal (i.e. VA) ERP waveform was compared with that of the summed unimodal (i.e. V + A–N) ERP waveform within each PD/ND interval separately for the high and low ABE groups, by conducting *repeated-measure* ANOVA with a single factor of ERP type [VA v.s. (V + A - N)] for each PD/ND component and separately for the two groups, in order to examine whether these cross-modal ERPs occur in both high and low ABE groups. More importantly, to investigate whether early cross-modal interactions underlie the individual difference in the magnitude of behavioral ABE, the amplitudes of each PD/ND component were then compared directly between the high and low ABE groups using *between-subject* ANOVA with a single factor of group (high v.s. low ABE group), respectively. All reported *p*-values for ANOVA results were corrected using the Greenhouse-Geisser method. It is noteworthy that (i) the main effect of the one-way between-subject ANOVA on PD/ND amplitudes (i.e. amplitudes of the bimodal-minus-unimodal difference waveforms) is equivalent to the interaction effect of the two-way ANOVA on ERP type (bimodal v.s. summed unimodal) and group (high v.s. low ABE group); (ii) our two-step analysis plan described above is equivalent to performing this two-way mixed ANOVA and then running paired *t*-tests to determine whether the bimodal and summed unimodal amplitudes differed in each group; (iii) analyzing amplitudes of the bimodal-minus-unimodal difference waveforms reduces the total number of statistical tests conducted, thereby controlling the family-wise error rate (Luck, 2014; Pierce, McDonald, & Green, 2018). Finally, a correlation analysis was performed across all 44 participants to further characterize the potential relationship between these cross-modal ERP amplitudes and the behavioral ABE magnitude.

## 3. Results

### 3.1. Behavioral results

For all 44 participants, the group mean percentage of bouncing percept in VA (i.e. audiovisual motion) condition was found to be

significantly higher than that in V (i.e. visual-only motion) condition [V: 4.2 ± 0.6% (mean ± *SE*); VA: 28.6 ± 4.1%; $F(1, 43) = 37.55$, $p < 0.0001$, $\eta_p^2 = 0.47$; Fig. 2A]. For the 30 participants whose ERP data were analyzed on a percept-related basis (see Percept-based ERP analysis section), the group mean percentage of bouncing percept was also significantly higher in VA condition (41.2 ± 4.3%) compared to V condition [5.4 ± 0.8%; $F(1, 29) = 63.91$, $p < 0.0001$, $\eta_p^2 = 0.69$; Fig. 2B]. These behavioral results replicated the classic audiovisual bounce-inducing effect (ABE) introduced by Sekuler et al. (1997). More importantly, the slow-moving velocity of the two visual disks used in the present study (see Stimuli and task section) led to an extremely low percentage of bouncing percept in V condition, which was in close agreement with observations in previous studies that also utilized slow velocity versions of streaming/bouncing visual motion (Bushara et al., 2003; Grassi & Casco, 2010, 2012; Watanabe & Shimojo, 2001b; Maniglia et al., 2012; Kawachi, 2016). This quite low percentage of bouncing percept in V condition indicated that almost all visual-only motion trials were perceived as streaming event. That is, the visual-only motion display designed in the current study was actually unambiguous. Therefore, the perception of the bouncing event in VA condition could be approximately considered as resulting solely from the ABE in the present study.

Besides, a 2 [group: high v.s. low ABE group (see Between-subject ERP analysis section for the division of the two groups)] × 2 (condition: V v.s. VA) two-way mixed ANOVA on the percentage of bouncing was conducted prior to the between-subject ERP analysis (see below). This ANOVA revealed a highly significant group × condition interaction [$F(1, 42) = 93.88$, $p < 0.0001$, $\eta_p^2 = 0.69$]. Specific contrasts (see Fig. 2C) showed that the percentage of bouncing was much higher for the high than low ABE group only in VA condition [high ABE group: 50.8 ± 4.3%; low ABE group: 6.4 ± 1.8%; $F(1, 42) = 91.88$, $p < 0.0001$, $\eta_p^2 = 0.69$] but not at all in V condition [high ABE group: 4.8 ± 0.8%; low ABE group: 3.7 ± 1.0%; $F(1, 42) = 0.67$, $p = 0.418$, $\eta_p^2 = 0.02$]. These results not only indicated that the difference in ABE magnitude between the high and low ABE groups mainly resulted from their difference in percentage of bouncing under VA condition, but also suggested that the two groups did not differ in their perceptual or decision criteria when perceiving the unimodal visual-only motion.

In addition, the false alarm rate in catch trials (i.e. the percentage of erroneous bouncing or streaming responses to the catch trials) was only 2.2 ± 0.3% on average, suggesting that our participants performed the task, as instructed, based veridically on their perceptual outcome after the coincidence event occurred, instead of simply relying on guesswork before that event occurred.

### 3.2. ERP results

#### 3.2.1. Variations of early cross-modal interactions underlie the occurrence of the ABE

For the ***VA_bouncing*** trials that reflected the unconfounded ABE trials (Fig. 3A), the first prominent cross-modal ERP was a positive difference component during 125–175 ms after sound onset with the largest amplitude over fronto-central electrodes (labeled as **PD170**). The PD170 was followed immediately by another positive difference that had a maximal amplitude over the bilateral parieto-occipital region within 180–200 ms interval (labeled as **PD190**). The last prominent cross-modal ERP was a large, broad negative difference within 220–280 ms time window, which was also largest over fronto-central scalp (labeled as **ND250**). To examine whether each of these cross-modal ERPs was present or absent on VA_bouncing trials, the significance of each component's amplitude was tested by repeated-measure ANOVA with a single factor of ERP type [bimodal ERP waveform VA_*bou* v.s. summed unimodal ERP waveform (V + A–N)], respectively. The results showed that both the fronto-central PD170 and occipital PD190 components were significant [PD170: $F(1, 29) = 7.47$, $p < 0.011$, $\eta_p^2 = 0.21$; PD190: $F(1, 29) = 11.35$, $p < 0.003$, $\eta_p^2 = 0.28$], with larger positive-going amplitudes for bimodal ERP than summed unimodal ERP waveform. The subsequent fronto-central ND250 component was also highly significant [$F(1, 29) = 53.08$, $p < 0.0001$, $\eta_p^2 = 0.65$], which was characterized as greater negative-going amplitude in bimodal ERP waveform than in summed unimodal ERP waveform. These results demonstrated that early cross-modal interactions were observed when bouncing event was perceived in VA condition. More importantly, given that VA_bouncing trials in the current study could be considered as being equivalent to ABE trials, these results thus also suggested that early cross-modal interactions occurred when the ABE was elicited.

For the ***VA_streaming*** trials on which the ABE was not elicited (Fig. 3B), instead, the earliest fronto-central **PD170** component did not show a significant amplitude at all [VA_*str* v.s. (V + A–N): $F(1, 29) = 0.18$, $p = 0.676$, $\eta_p^2 = 0.006$], which replicated the recent finding of Zhao et al. (2018). Furthermore, the subsequent occipital **PD190** component was also found to be nonsignificant [$F(1, 29) = 2.09$, $p = 0.159$, $\eta_p^2 = 0.07$]. The later fronto-central **ND250** component, however, was still highly significant on VA_bouncing trials [$F(1, 29) = 82.36$, $p < 0.0001$, $\eta_p^2 = 0.74$], with substantially larger negative-going amplitude in bimodal ERP waveform than in summed unimodal ERP waveform. These results indicated that although the relatively late cross-modal interaction manifested by ND250 component was also elicited on the non-ABE trials, cross-modal neural activities at early processing stage (earlier than 200 ms after sound onset) were actually absent.

To further examine whether the variations of early cross-modal interactions underlie the occurrence of the ABE, the amplitude of each PD/ND component was then compared directly (see Fig. 4) using repeated-measure ANOVA with a single factor of perception (CMdiff_bou v.s. CMdiff_str). The results showed that the fronto-central **PD170** amplitude was significantly larger [$F(1, 29) = 7.01$, $p < 0.013$, $\eta_p^2 = 0.20$] on CMdiff_bou waveform [0.81 ± 0.30 μV (*M* ± *SE*)] than on CMdiff_str waveform (0.12 ± 0.29 μV). Moreover, the following occipital **PD190** component was also found to be substantially greater [$F(1, 29) = 4.93$, $p < 0.035$, $\eta_p^2 = 0.15$] for CMdiff_bou (0.77 ± 0.23 μV) than CMdiff_str waveform (0.28 ± 0.19 μV). The subsequent fronto-central negativity **ND250**, however, *did not* differ significantly in amplitude [$F(1, 29) = 0.72$, $p = 0.403$, $\eta_p^2 = 0.02$] between CMdiff_bou (-3.53 ± 0.48 μV) and CMdiff_str (-3.90 ± 0.43 μV) waveforms. In general, these results were highly consistent with the ERP results recently reported by Zhao et al. (2018). Most importantly, since bouncing trials in VA condition in the present study actually denote trials on which the ABE occurred, these results above thus provide unconfounded evidence for the hypothesis that early cross-modal interactions underlie the occurrence of the ABE phenomenon.

#### 3.2.2. Cross-modal interaction at later stage underlies the individual difference in tendency to perceive the ABE

For the ***high ABE group*** (Fig. 5A), the significance of each cross-modal difference component was tested using *repeated-measure* ANOVA with a single factor of ERP type [bimodal ERP waveform VA v.s. summed unimodal ERP waveform (V + A–N)], respectively. The results revealed that both the fronto-central **PD170** and occipital **PD190** components were significant [PD170: $F(1, 21) = 4.93$, $p < 0.038$, $\eta_p^2 = 0.19$; PD190: $F(1, 21) = 7.23$, $p < 0.014$, $\eta_p^2 = 0.26$], with larger positive-going amplitudes for bimodal ERP than summed unimodal ERP waveform. The later fronto-central **ND250** component was also found to be significant [$F(1, 21) = 54.44$, $p < 0.0001$, $\eta_p^2 = 0.72$], which resulted from greater negative-going amplitude on bimodal ERP waveform than on summed unimodal ERP waveform.

For the ***low ABE group*** (Fig. 5B), the one-way *repeated-measure* ANOVA for the fronto-central **PD170** component did not show a significant main effect of ERP type [$F(1, 21) = 0.97$, $p = 0.337$, $\eta_p^2 = 0.04$]. However, the subsequent occipital **PD190** component was found to be still significant in the low ABE group [$F(1, 21) = 6.96$, $p <$

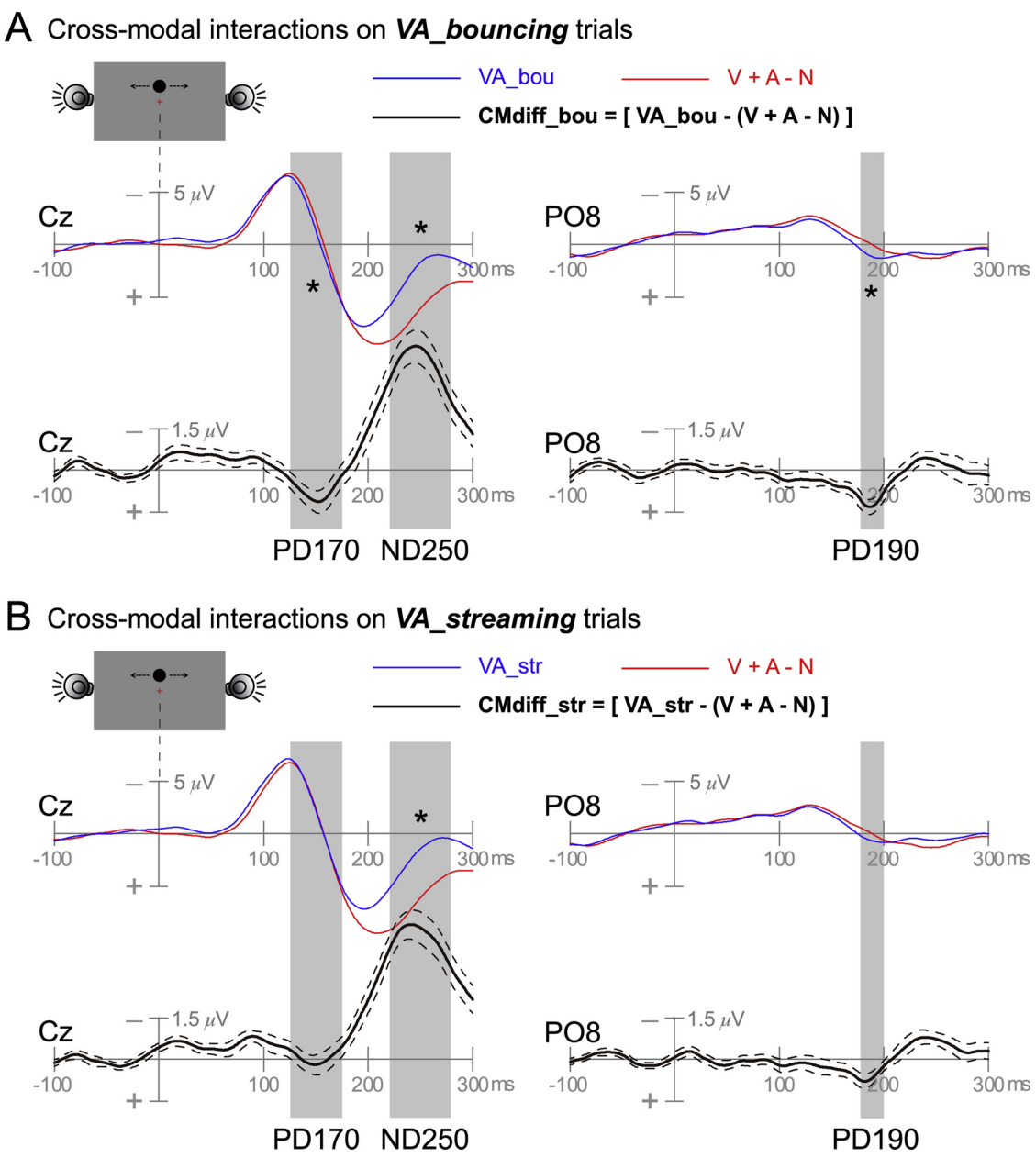


**Fig. 3.** (**A**) Cross-modal interactions on ***VA_bouncing*** trials (i.e. the audiovisual trials on which bouncing event was perceived) for the 30 participants whose ERP data were analyzed on a percept-related basis: Grand-averaged bimodal ERP waveform elicited on VA_bouncing trials [blue line, time-locked to the sound onset (i.e. the frame 11 onset), the same below], summed unimodal ERP waveforms elicited in V and A conditions (red line), as well as the calculated CMdiff_bou waveform reflecting cross-modal interactions when perceiving bouncing event in VA condition (black line; dotted lines indicate ± 1 SEM) are shown from Cz and PO8 electrode sites. The shaded areas depict the time windows within which the cross-modal difference ERP components [the fronto-central PD170 (125–175 ms after sound onset), the occipital PD190 (180–200 ms), the fronto-central ND250 (220–280 ms)] were measured, respectively. The symbol '*' denotes the *occurrence* of significant cross-modal interaction ($p < 0.05$). (**B**) Same as Fig. 3A but for cross-modal interactions on ***VA_streaming*** trials (i.e. the audiovisual trials on which subjects still reported the streaming percept). Note that early cross-modal interactions indexed by the PD170 and PD190 components were present on VA_bouncing trials but absent on VA_streaming trials.

0.016, $\eta_p^2 = 0.25$], which was manifested by larger positive amplitude on bimodal ERP waveform than on summed unimodal ERP waveform. Similarly, the relatively late fronto-central negativity **ND250** was also significant [$F(1, 21) = 72.41$, $p < 0.0001$, $\eta_p^2 = 0.78$], with greater negative-going amplitude for bimodal ERP than summed unimodal ERP waveform.

In order to examine directly whether cross-modal interactions contribute to the individual difference in predisposition to experience the ABE, the amplitudes of each PD/ND component in CMdiff waveform were then contrasted directly between high and low ABE groups (see Fig. 6A & B). The results showed that, although the cross-modal interaction revealed by fronto-central **PD170** was only significant for high ABE group [0.71 ± 0.32 μV ($M \pm SE$)] but not for low ABE group (0.29 ± 0.30 μV), no significant group difference was actually found for the PD170 component [$F(1, 42) = 0.90$, $p = 0.349$, $\eta_p^2 = 0.02$]. Following the PD170, the amplitude of the occipital **PD190** component did not differ significantly between high ABE group and low ABE group either [$F(1, 42) = 0.001$, $p = 0.978$, $\eta_p^2 < 0.0001$; high ABE group: 0.62 ± 0.23 μV; low ABE group: 0.61 ± 0.23 μV]. However, a significant group difference was found for the relatively late, fronto-central **ND250** component [$F(1, 42) = 7.23$, $p < 0.011$, $\eta_p^2 = 0.15$], with prominently larger ND250 amplitude for high ABE group

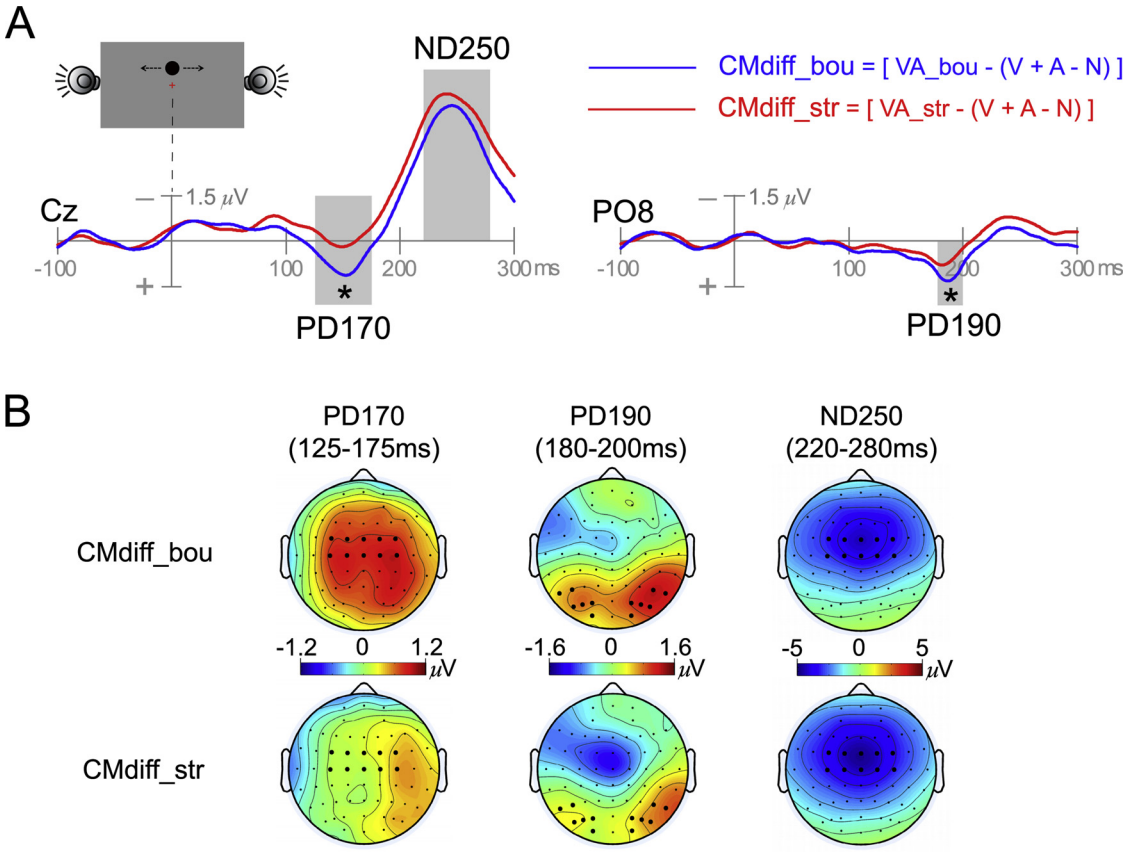


**Fig. 4.** Direct comparisons of cross-modal interactions between VA_bouncing (CMdiff_bou) and VA_streaming trials (CMdiff_str) for the 30 participants whose ERP data were analyzed on a percept-related basis: (**A**) Comparisons of the amplitudes of cross-modal ERP components between CMdiff_bou and CMdiff_str difference waveforms are shown from Cz and PO8 electrodes. The symbol '*' here indicates significant *difference* in cross-modal interaction between CMdiff_bou and CMdiff_str waveforms ($p < 0.05$). (**B**) Scalp topographies of these cross-modal ERP components in the CMdiff_bou and CMdiff_str waveforms are shown in top view. The bold dots on each scalp topography depict the electrode sites over which the corresponding ERP component was quantified (for details, see percept-based ERP analysis section). Note that early cross-modal interactions indexed by the PD170 and PD190 components were significantly larger in CMdiff_bou than in CMdiff_str waveform.

(-4.02 ± 0.54 μV) than for low ABE group (-2.37 ± 0.28 μV; see Fig. 6A & B). Of note, the two groups of subjects did not differ in amplitude during the ND250 interval (220–280 ms) on any of the unisensory ERP waveforms that were used to calculate the cross-modal difference waveform [V condition: $F(1, 42) = 0.008$, $p = 0.927$, $\eta_p^2 < 0.0003$; A condition: $F(1, 42) = 2.80$, $p = 0.101$, $\eta_p^2 = 0.06$; N condition: $F(1, 42) = 0.001$, $p = 0.970$, $\eta_p^2 < 0.0001$].

Lastly, a correlation analysis was performed across all 44 participants to further examine whether the between-subject difference in ND250 amplitude could account for the inter-individual variability in the behavioral ABE magnitude. Indeed, a significant negative correlation was found between the fronto-central **ND250** amplitude and the ABE magnitude [$r(42) = -0.34$, $p < 0.026$; see Fig. 6C], indicating that subjects who had a larger ND250 amplitude physiologically (i.e. more negative amplitude) tended to show a higher ABE magnitude behaviorally. In contrast, earlier cross-modal neural activities manifested by the fronto-central **PD170** and occipital **PD190** amplitudes were not correlated with the behavioral ABE magnitude [PD170: $r(42) = 0.19$, $p = 0.225$; PD190: $r(42) = 0.10$, $p = 0.514$; see Fig. 6C]. Generally, these results above demonstrated that cross-modal interaction occurring at relatively late processing stage underlies the inter-individual variability in predisposition to perceive the ABE phenomenon.

## 4. Discussion

Previous neuroscience studies (Bushara et al., 2003; Hipp et al., 2011; Zhao et al., 2018; Zvyagintsev et al., 2011) mainly adopted the method of comparing the audiovisual bouncing with streaming trials to explore the neural mechanisms of the ABE introduced by Sekuler et al. (1997). However, as the bouncing event could be also perceived in the visual-only motion display (although occasionally) in these studies (except for Bushara et al., 2003), the audiovisual bouncing trials consisted of not only the trials on which the ABE occurred but also certain bouncing trials on which the ABE did not necessarily occur (see Introduction). Therefore, the present ERP study refined the experimental paradigm by designing an unambiguous version of visual streaming/bouncing display in which almost all visual-only trials were perceived as streaming, thus the bouncing responses in audiovisual trials could be considered as resulting solely from the ABE. The behavioral results in the present study showed a reliable ABE effect, which was characterized as much higher percentage of bouncing percept in VA condition than in V condition. More importantly, the slow motion speed of the two visual disks used in the present study (for details, see Stimuli and task section), as expected, led to few bouncing responses in V condition (only 4% on average over all subjects), which is highly consistent with results in previous ABE studies that also utilized slow velocity versions of streaming/bouncing visual motion (Bushara et al., 2003; Grassi & Casco, 2010, 2012; Watanabe & Shimojo, 2001b; Maniglia et al., 2012; Kawachi, 2016). This extremely low incidence of bouncing percept in V condition confirmed that the visual-only motion display designed in the present study was unambiguous and suggested that the experience of bouncing percept under VA condition could be considered as stemming from the ABE.

The behavioral results mentioned above indicate that the present bouncing responses in VA condition (labeled as *VA_bouncing* trials)

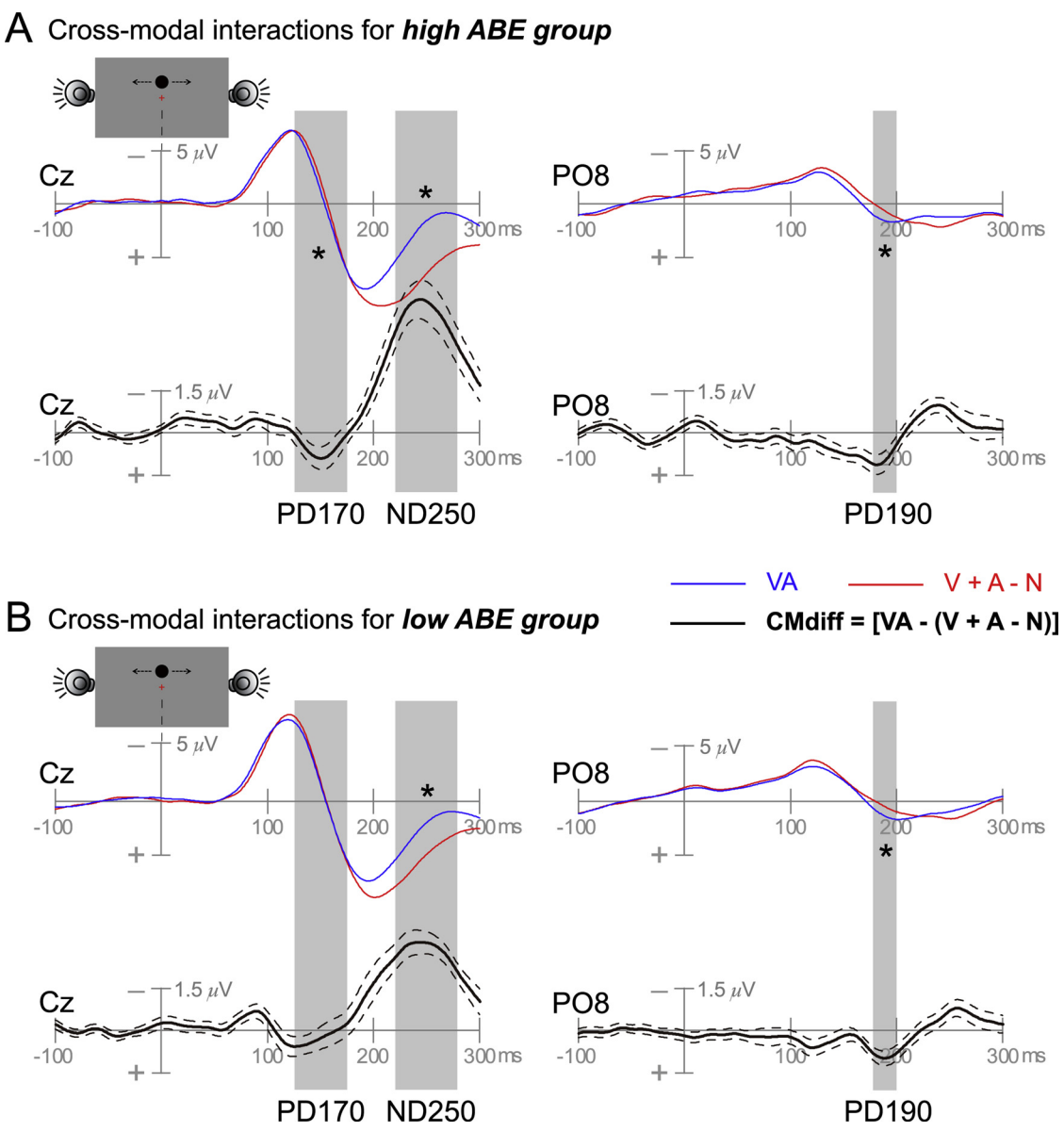


**Fig. 5.** Cross-modal interactions for subjects who had a higher behavioral ABE magnitude (the ***high ABE group***, $n = 22$, **A**) and for those who showed a lower or no ABE magnitude (the ***low ABE group***, $n = 22$, **B**): Bimodal ERP waveform elicited in VA condition (blue line, without separation between VA_bouncing and VA_streaming trials), summed unimodal ERP waveforms elicited in V and A conditions (red line), as well as the calculated CMdiff waveform reflecting cross-modal interactions (black line; dotted lines indicate ± 1 SEM) were grand-averaged separately for the high and low ABE groups. Example waveforms shown were recorded from Cz and PO8 electrodes. The time intervals (shaded areas) selected for measuring cross-modal ERPs (i.e. PD170, PD190 and ND250) in the CMdiff waveform were identical to those used for percept-based analysis (see Fig. 3A). The symbol '*' denotes the *occurrence* of significant cross-modal interaction ($p < 0.05$).

could represent unconfounded ABE trials (i.e. trials on which the ABE occurred), thus are more suitable for being compared with *VA_streaming* trials (i.e. trials on which the ABE did not occur) to examine whether early cross-modal interactions underlie the occurrence of the ABE. Based on this refinement, the present percept-based ERP analysis showed that two early cross-modal ERP components, the fronto-central PD170 (125–175 ms after sound onset) and the occipital PD190 (180–200 ms), were significantly larger for VA_bouncing than VA_streaming trials (Fig. 4). In particular, both the PD170 and PD190 components were actually *absent* on VA_streaming trials where the ABE was not induced (Fig. 3B). These ERP results are in close agreement with the findings recently reported by Zhao et al. (2018). Most importantly, given that VA_bouncing trials in the present study represent unconfounded ABE trials, these results thus provide more direct and precise evidence for the hypothesis that early cross-modal interactions underlie the occurrence of the ABE phenomenon.

The fronto-central positive difference PD170 has been found in many previous multisensory studies using the method of calculating cross-modal difference wave (e.g. Giard & Peronnet, 1999; Fort et al., 2002; Molholm et al., 2002; Talsma & Woldorff, 2005; Mishra et al., 2007, 2008, 2010, their PD180; Vidal, Giard, Roux, Barthélémy, & Bruneau, 2008; Zhao et al., 2018, their PD170). The amplitude of this cross-modal positivity has been found to correlate positively with sound-induced segmenting of visual inputs on a percept-related basis (Mishra et al., 2008). It is noteworthy that classic ABE studies (Grassi & Casco, 2009, their Exp. 5 & 6; Grassi & Casco, 2012) have shown that if the two visual disks could only *overlap partially* with each other and then moved apart (which can be considered as separated visual inputs), bouncing responses would be markedly increased relative to when the two visual disks could *overlap completely* (which can be considered as fused visual inputs). More importantly, recent studies found that presenting sounds at the coincident moment of the two disks induced a

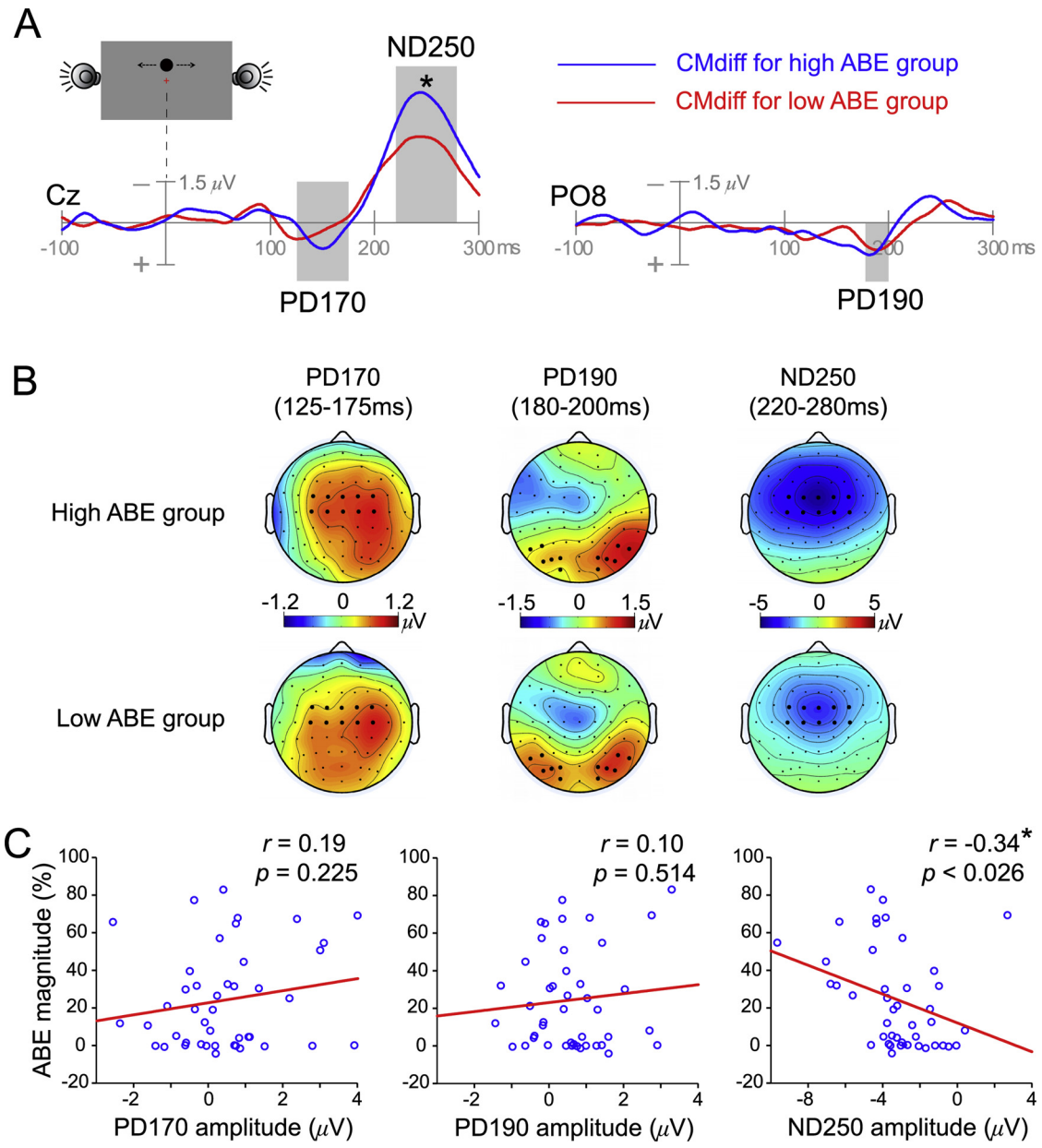


**Fig. 6.** (**A**) Direct comparisons of cross-modal ERP components in CMdiff waveform between the high and low ABE groups are shown from Cz and PO8 electrodes. The symbol '*' here indicates significant *difference* in cross-modal interaction between the high and low ABE groups ($p < 0.05$). (**B**) Scalp topographies of these cross-modal ERPs in the high and low ABE groups are shown in top view. Note that a significant group difference was found only for the relatively late ND250 component. (**C**) Scatter plots showing correlations between the behavioral ABE magnitude (i.e. the difference in percentage of bouncing percept between VA and V conditions) and the amplitudes of PD170, PD190 and ND250 components across all 44 participants. A significant correlation with the behavioral ABE magnitude was found only for the fronto-central ND250 amplitude, indicating subjects with larger ND250 amplitudes have higher tendencies to experience the ABE.

strong "non-overlapping" bias even when the two disks could overlap completely (Kawachi, 2016; Meyerhoff & Scholl, 2018), which was thought to be a proximate cause driving the ABE (for details, see Meyerhoff & Scholl, 2018). Combining these electrophysiological and behavioral findings, larger PD170 amplitude on VA_bouncing than VA_streaming trials observed in the present study might indicate that: the coincident sound elicited much stronger subjective impression of only partial overlap of the two visual disks (i.e. stronger segmenting of the two disks at the moment of coincidence) on some trials than others, and the trials with stronger subjective impression of incomplete overlap were thus perceived as bouncing event (otherwise, still streaming percept).

If the coincident sound indeed promoted stronger visual perception of incomplete overlap on the subsequent bouncing than streaming trials, we would also expect different cross-modal neural activity over the visual region between VA _bouncing and VA_streaming trials. As expected, the cross-modal positivity PD190 (180–200 ms after sound onset) over occipital scalp was found to be greater for VA_bouncing than VA_streaming trials in the present study (Fig. 4). This occipital positive difference component has also been reported in a large number of past ERP studies on cross-modal interactions (e.g., Giard & Peronnet, 1999; Molholm et al., 2002; Teder-Sälejärvi et al., 2002, 2005, their P190; Wu, Li, Bai, & Touge, 2009; Yang et al., 2013; Gao et al., 2014; Zhao et al., 2018, their PD190). As the latency and scalp topography of this cross-modal positivity were highly similar to those of the visual N1 component, it was typically considered as originating from smaller visual N1 amplitude elicited by bimodal stimuli (AV) than the sum of unimodal stimuli (A + V), which indicated the influence of auditory

inputs on the processing in visual cortex (Giard & Peronnet, 1999; Molholm et al., 2002). This interpretation was further supported by findings that the neural generators of this occipital cross-modal positivity were localized to the ventral extrastriate visual cortex (2005, Teder-Sälejärvi et al., 2002). Therefore, the present finding of larger occipital PD190 amplitude on VA_bouncing than VA_streaming trials is exactly what would be expected if the coincident sound exerted a stronger influence on the processing of the motion signals in the visual cortex, or more specifically, induced stronger visual impression of incomplete overlap, for the subsequent bouncing than streaming trials. Even without these assumptions mentioned above, the variations of the fronto-central PD170 and occipital PD190 amplitudes were closely associated with the occurrence of the ABE, which was enough to demonstrate the ABE phenomenon originates, at least in part, from audiovisual cross-modal interactions elicited at relatively early processing stage.

A between-subject ERP analysis was also conducted to investigate whether cross-modal interactions also underlie the inter-individual variability in predisposition to experience the ABE (i.e. the between-subject difference in the ABE magnitude). To this end, participants were divided, by a median split of the ABE magnitude, into two groups (22 in each) that had a higher ABE magnitude (the *high ABE group*) and lower or even no ABE magnitude (the *low ABE group*). The cross-modal difference waveforms were then compared between the high and low ABE groups. This between-subject ERP analysis revealed that earlier cross-modal interactions indexed by the PD170 and PD190 components did not differ between the high and low ABE groups, indicating these earlier cross-modal interactions underlying the occurrence of the ABE are *necessary but not sufficient* for eventually determining the ABE magnitude. Instead, the relatively late, fronto-central negativity ND250 (220–280 ms after sound onset) was found to be much greater in the high than low ABE group (Fig. 6A & B). Additional correlation analysis further showed that subjects with larger ND250 amplitudes had higher tendencies to experience the ABE. These results demonstrate that cross-modal interaction occurring at relatively late processing stage might be responsible for the inter-individual variability in predisposition to perceive the ABE phenomenon.

A broad fronto-central negativity closely resembling the present ND250 component has been consistently observed in a series of cross-modal ERP studies that calculated cross-modal difference waveforms (e.g. Teder-Sälejärvi et al., 2002, 2005, their N260; Talsma & Woldorff, 2005; Mishra et al., 2007, their ND270; Bonath et al., 2007, their N260; Mishra et al., 2008, their ND240; Wu et al., 2009; Mishra, Martinez, & Hillyard, 2010, their ND250/ND240; Li et al., 2010; Gao et al., 2014; Zhao et al., 2018, their ND250). This cross-modal negative deflection was thought to reflect a general aspect of cross-modal interaction (Mishra et al., 2007, 2010), or more specifically, reflect a default mode of multisensory integration by which the visual and auditory inputs would be bound into a coherent object (but at later processing stage) based on their temporal and/or spatial co-occurrence even when only the visual modality was attended (Busse, Roberts, Crist, Weissman, & Woldorff, 2005; Talsma et al., 2007), and regardless of the strength of associations in meaning between the visual and auditory inputs (Fiebelkorn, Foxe, & Molholm, 2010; Vroomen & Stekelenburg, 2009).

According to the hypothesis above, we would expect no difference in ND250 amplitude between VA_bouncing trials (on which stronger association between the coincidence of the two visual disks and the coincident sound seemed to be perceived) and VA_streaming trials (on which weaker association seemed to be perceived). Indeed, the present percept-based analysis of the ND250 component *did not* find significant amplitude difference between VA_bouncing and VA_streaming trials (see Fig. 4). Therefore, the present inter-individual difference on ND250 amplitude might reflect inherent differences in multisensory binding tendency between the high and low ABE groups of subjects, and this tendency to bind multisensory features at relatively late processing stage seems to be also *necessary but not sufficient* for triggering the ABE (indexed by diminished ND250 in the low ABE group but no ND250 difference between the ABE versus non-ABE trials). Alternatively, it might be argued that the present group difference in ND250 was merely the consequence of difference in response criterion or decision-making process given its relatively late latency and fronto-central scalp distribution. However, the neural generators of this broad negativity have been localized to either the vicinity of auditory cortex (Bonath et al., 2007; Teder-Sälejärvi et al., 2002) or the polysensory superior temporal cortex (2008, Mishra et al., 2007, 2010; Teder-Sälejärvi et al., 2005), but not the frontal region. Accordingly, the present group difference in activation over sensory-related brain areas seems less likely to be accounted for by the difference in response criterion or decision-making process between the two groups.

A prior EEG study focusing on oscillatory synchronization (Hipp et al., 2011) has found that the high gamma band (about 80 Hz) synchronization across central and temporal regions could also account for the individual difference in predisposition to perceive the ABE. However, this gamma rhythm synchronization was found to be also different between bouncing and streaming trials in VA condition, which was inconsistent with the present result of no ND250 difference between the ABE versus non-ABE trials. Besides, this ABE-related gamma band coherence was found to be prominent even about 100 ms *before* the sound onset (see Hipp et al., 2011), whereas the present ND250 component was evident about 200 ms *after* the presentation of the sound. Therefore, the present ND250 seems less likely to be the averaged ERP counterpart of the gamma rhythm coherence reported by Hipp et al. (2011), indicating they might reflect distinct psychophysiological process. Additional study might be needed to further distinguish the specific roles the ND250 component and the gamma band synchronization played in the tendency to perceive the ABE phenomenon.

It is also worth mentioning that a recent behavioral study investigated the ABE among healthy old subjects and found that the ABE magnitude was substantially lower for the elderly than the younger after ruling out the contributions of age-related changes in vision and audition, suggesting weakened inter-modal integration with aging (Roudaia et al., 2013). Given that the ND250 component was thought to reflect general audiovisual binding tendency (Fiebelkorn et al., 2010; Mishra et al., 2007, 2010; Vroomen & Stekelenburg, 2009) and was found to be correlated with the inter-individual variability in predisposition to perceive the ABE, it is interesting for future study to examine whether the cross-modal interaction revealed by ND250 also underlies the age-related reduction in ABE magnitude reported by Roudaia and colleagues.

## 5. Conclusion

In summary, the present ERP study investigated the neural activities associated with early cross-modal interactions underlie the audiovisual bounce-inducing effect (ABE) in a refined task. The proximate triggers for the occurrence of ABE appear to be early cross-modal interactions over the fronto-central region (PD170, 125–175 ms after sound onset) and occipital scalp (PD190, 180–200 ms), which were much larger when the ABE occurred but absent when the ABE did not occur. Furthermore, subjects who were disposed to perceive the ABE more frequently showed a much larger cross-modal interaction at later latency over fronto-central region (ND250, 220–280 ms), but this cross-modal negativity did not differ between the ABE verse non-ABE trials in the within-subject analysis, indicating a strong audiovisual binding tendency in general might be necessary but not sufficient for triggering the ABE. Based on these findings, the present study proposes that the ABE is generated, at least in part, as a consequence of the rapid interplay between the variations of early cross-modal interactions and the general multisensory integration predisposition at an individual level. A challenge for the future is to determine the exact anatomical pathways through which these cross-modal neural activities interact to produce the ABE phenomenon.

## Author contributions

S.Z. and W.F. designed the research; S.Z. and Y.W. performed the research; S.Z., Y.W. and W.F. analyzed the data; S.Z., C.F., and W.F. wrote the paper.

## Declaration of Competing Interest

None.

## Acknowledgments

This work was supported by the National Natural Science Foundation of China (grant numbers 31400868 and 31771200 to W.F.F.).

## References

Berger, C. C., & Ehrsson, H. H. (2013). Mental imagery changes multisensory perception. *Current Biology, 23*(14), 1367–1372.

Berger, C. C., & Ehrsson, H. H. (2017). The content of imagined sounds changes visual motion perception in the cross-bounce illusion. *Scientific Reports, 7*, 40123. https://doi.org/10.1038/srep40123.

Bertelson, P., & Aschersleben, G. (1998). Automatic visual bias of perceived auditory location. *Psychonomic Bulletin & Review, 5*(3), 482–489.

Bertenthal, B. I., Banton, T., & Bradbury, A. (1993). Directional bias in the perception of translating patterns. *Perception, 22*(2), 193–207.

Bonath, B., Noesselt, T., Martinez, A., Mishra, J., Schwiecker, K., Heinze, H. J., et al. (2007). Neural basis of the ventriloquist illusion. *Current Biology, 17*(19), 1697–1703.

Bushara, K. O., Hanakawa, T., Immisch, I., Toma, K., Kansaku, K., & Hallett, M. (2003). Neural correlates of cross-modal binding. *Nature Neuroscience, 6*(2), 190–195.

Busse, L., Roberts, K. C., Crist, R. E., Weissman, D. H., & Woldorff, M. G. (2005). The spread of attention across modalities and space in a multisensory object. *Proceedings of the National Academy of Sciences, 102*(51), 18751–18756.

Capizzi, M., Correa, Á., & Sanabria, D. (2013). Temporal orienting of attention is interfered by concurrent working memory updating. *Neuropsychologia, 51*(2), 326–339.

Cappe, C., Thelen, A., Romei, V., Thut, G., & Murray, M. M. (2012). Looming signals reveal synergistic principles of multisensory integration. *Journal of Neuroscience, 32*(4), 1171–1182.

Cappe, C., Thut, G., Romei, V., & Murray, M. M. (2010). Auditory-visual multisensory interactions in humans: Timing, topography, directionality, and sources. *Journal of Neuroscience, 30*(38), 12572–12580.

Donohue, S. E., Green, J. J., & Woldorff, M. G. (2015). The effects of attention on the temporal integration of multisensory stimuli. *Frontiers in Integrative Neuroscience, 9*, 32. https://doi.org/10.3389/fnint.2015.00032.

Driver, J., & Noesselt, T. (2008). Multisensory interplay reveals crossmodal influences on 'sensory-specific' brain regions, neural responses, and judgments. *Neuron, 57*(1), 11–23.

Dufour, A., Touzalin, P., Moessinger, M., Brochard, R., & Després, O. (2008). Visual motion disambiguation by a subliminal sound. *Consciousness and Cognition, 17*(3), 790–797.

Fiebelkorn, I. C., Foxe, J. J., & Molholm, S. (2010). Dual mechanisms for the cross-sensory spread of attention: how much do learned associations matter? *Cerebral Cortex, 20*(1), 109–120.

Fort, A., Delpuech, C., Pernier, J., & Giard, M. H. (2002). Dynamics of cortico-subcortical cross-modal operations involved in audio-visual object detection in humans. *Cerebral Cortex, 12*(10), 1031–1039.

Fujisaki, W., Shimojo, S., Kashino, M., & Nishida, S. (2004). Recalibration of audiovisual simultaneity. *Nature Neuroscience, 7*(7), 773–778.

Gao, Y., Li, Q., Yang, W., Yang, J., Tang, X., & Wu, J. (2014). Effects of ipsilateral and bilateral auditory stimuli on audiovisual integration: A behavioral and event-related potential study. *NeuroReport, 25*(9), 668–675.

Giard, M. H., & Peronnet, F. (1999). Auditory-visual integration during multimodal object recognition in humans: A behavioral and electrophysiological study. *Journal of Cognitive Neuroscience, 11*(5), 473–490.

Gobara, A., Yoshimura, N., & Yamada, Y. (2018). Arousing emoticons edit stream/bounce perception of objects moving past each other. *Scientific Reports, 8*, 5752. https://doi.org/10.1038/s41598-018-23973-4.

Gondan, M., & Röder, B. (2006). A new method for detecting interactions between the senses in event-related potentials. *Brain Research, 1073–1074*(1), 389–397.

Grassi, M., & Casco, C. (2009). Audiovisual bounce-inducing effect: Attention alone does not explain why the discs are bouncing. *Journal of Experimental Psychology Human Perception and Performance, 35*(1), 235–243.

Grassi, M., & Casco, C. (2010). Audiovisual bounce-inducing effect: When sound congruence affects grouping in vision. *Attention, Perception & Psychophysics, 72*(2), 378–386.

Grassi, M., & Casco, C. (2012). Revealing the origin of the audiovisual bounce-inducing effect. *Seeing and Perceiving, 25*(2), 223–233.

Grove, P. M., Ashton, J., Kawachi, Y., & Sakurai, K. (2012). Auditory transients do not affect visual sensitivity in discriminating between objective streaming and bouncing events. *Journal of Vision, 12*(8), 1–11.

Grove, P. M., Robertson, C., & Harris, L. R. (2016). Disambiguating the stream/bounce illusion with inference. *Multisensory Research, 29*, 453–464.

Grove, P. M., & Sakurai, K. (2009). Auditory induced bounce perception persists as the probability of a motion reversal is reduced. *Perception, 38*(7), 951–965.

Hipp, J. F., Engel, A. K., & Siegel, M. (2011). Oscillatory synchronization in large-scale cortical networks predicts perception. *Neuron, 69*(2), 387–396.

Kawabe, T., & Miura, K. (2006). Effects of the orientation of moving objects on the perception of streaming/bouncing motion displays. *Perception & Psychophysics, 68*(5), 750–758.

Kawachi, Y. (2016). Visual mislocalization of moving objects in an audiovisual event. *PloS One, 11*(4), e0154147. https://doi.org/10.1371/journal.pone.0154147.

Li, Q., Wu, J., & Touge, T. (2010). Audiovisual interaction enhances auditory detection in late stage: An event-related potential study. *NeuroReport, 21*(3), 173–178.

Luck, S. J. (2014). *An introduction to the event-related potential technique.* Cambridge, MA: MIT press.

Maniglia, M., Grassi, M., Casco, C., & Campana, G. (2012). The origin of the audiovisual bounce-inducing effect: A TMS study. *Neuropsychologia, 50*(7), 1478–1482.

Metzger, W. (1934). Beobachtungen über phänomenale identität. *Psychologische Forschung, 19*(1), 1–60.

Meyerhoff, H. S., Merz, S., & Frings, C. (2018). Tactile stimulation disambiguates the perception of visual motion paths. *Psychonomic Bulletin & Review*. https://doi.org/10.3758/s13423-018-1467-0.

Meyerhoff, H. S., & Scholl, B. (2018). Auditory-induced bouncing is a perceptual (rather than a cognitive) phenomenon: Evidence from illusory crescents. *Cognition, 170*, 88–94.

Mishra, J., Martinez, A., & Hillyard, S. A. (2008). Cortical processes underlying sound-induced flash fusion. *Brain Research, 1242*(4), 102–115.

Mishra, J., Martinez, A., & Hillyard, S. A. (2010). Effect of attention on early cortical processes associated with the sound-induced extra flash illusion. *Journal of Cognitive Neuroscience, 22*(8), 1714–1729.

Mishra, J., Martinez, A., Sejnowski, T., & Hillyard, S. A. (2007). Early cross-modal interactions in auditory and visual cortex underlie a sound-induced visual illusion. *Journal of Neuroscience, 27*(15), 4120–4131.

Molholm, S., Ritter, W., Murray, M. M., Javitt, D. C., Schroeder, C. E., & Foxe, J. J. (2002). Multisensory auditory-visual interactions during early sensory processing in humans: A high-density electrical mapping study. *Cognitive Brain Research, 14*(1), 115–128.

Parise, C. V., & Ernst, M. O. (2017). Noise, multisensory integration, and previous response in perceptual disambiguation. *PLoS Computational Biology, 13*(7), e1005546. https://doi.org/10.1371/journal.pcbi.1005546.

Pierce, A. M., McDonald, J. J., & Green, J. J. (2018). Electrophysiological evidence of an attentional bias in crossmodal inhibition of return. *Neuropsychologia, 114*, 11–18.

Remijn, G. B., Ito, H., & Nakajima, Y. (2004). Audiovisual integration: An investigation of the 'streaming-bouncing' phenomenon. *Journal of Physiological Anthropology and Applied Human Science, 23*(6), 243–247.

Roudaia, E., Sekuler, A. B., Bennett, P. J., & Sekuler, R. (2013). Aging and audio-visual and multi-cue integration in motion. *Frontiers in Psychology, 4*, 267. https://doi.org/10.3389/fpsyg.2013.0026.

Sanabria, D., Correa, Á, Lupiáñez, J., & Spence, C. (2004). Bouncing or streaming? Exploring the influence of auditory cues on the interpretation of ambiguous visual motion. *Experimental Brain Research, 157*(4), 537–541.

Scheier, C., Lewkowicz, D. J., & Shimojo, S. (2003). Sound induces perceptual reorganization of an ambiguous motion display in human infants. *Developmental Science, 6*(3), 233–244.

Sekuler, A. B., & Sekuler, R. (1999). Collisions between moving visual targets: what controls alternative ways of seeing an ambiguous display? *Perception, 28*(4), 415–432.

Sekuler, R., Sekuler, A. B., & Lau, R. (1997). Sound alters visual motion perception. *Nature, 385*, 308.

Senkowski, D., Saint-Amour, D., Höfle, M., & Foxe, J. J. (2011). Multisensory interactions in early evoked brain activity follow the principle of inverse effectiveness. *NeuroImage, 56*, 2200–2208.

Shimojo, S., & Shams, L. (2001). Sensory modalities are not separate modalities: Plasticity and interactions. *Current Opinion in Neurobiology, 11*(4), 505–509.

Talsma, D., & Woldorff, M. G. (2005). Selective attention and multisensory integration: Multiple phases of effects on the evoked brain activity. *Journal of Cognitive Neuroscience, 17*(7), 1098–1114.

Talsma, D., Doty, T. J., & Woldorff, M. G. (2007). Selective attention and audiovisual integration: is attending to both modalities a prerequisite for early integration? *Cerebral Cortex, 17*(3), 679–690.

Teder-Sälejärvi, W. A., Di Russo, F., McDonald, J. J., & Hillyard, S. A. (2005). Effects of spatial congruity on audio-visual multimodal integration. *Journal of Cognitive Neuroscience, 17*(9), 1396–1409.

Teder-Sälejärvi, W. A., McDonald, J. J., Di Russo, F., & Hillyard, S. A. (2002). An analysis of audio-visual crossmodal integration by means of event-related potential (ERP) recordings. *Cognitive Brain Research, 14*(1), 106–114.

Van der Burg, E., Talsma, D., Olivers, C. N. L., Hickey, C., & Theeuwes, J. (2011). Early multisensory interactions affect the competition among multiple visual objects. *NeuroImage, 55*, 1208–1218.

Vidal, J., Giard, M. H., Roux, S., Barthélémy, C., & Bruneau, N. (2008). Cross-modal processing of auditory-visual stimuli in a no-task paradigm: A topographic event-related potential study. *Clinical Neurophysiology, 119*(4), 763–771.

Vroomen, J., & Stekelenburg, J. J. (2009). Visual anticipatory information modulates multisensory interactions of artificial audiovisual stimuli. *Journal of Cognitive Neuroscience, 22*(7), 1583–1596.

Walter, W. G., Cooper, R., Aldridge, V. J., McCallum, W. C., & Winter, A. L. (1964).

Contingent negative variation: An electric sign of sensori-motor association and expectancy in the human brain. *Nature, 203*(4), 380–384.
Watanabe, K., & Shimojo, S. (1998). Attentional modulation in perception of visual motion events. *Perception, 27*, 1041–1054.
Watanabe, K., & Shimojo, S. (2001a). Postcoincidence trajectory duration affects motion event perception. *Perception & Psychophysics, 63*(1), 16–28.
Watanabe, K., & Shimojo, S. (2001b). When sound affects vision: Effects of auditory grouping on visual motion perception. *Psychological Science, 12*(2), 109–116.
Wu, J., Li, Q., Bai, O., & Touge, T. (2009). Multisensory interactions elicited by audio-visual stimuli presented peripherally in a visual attention task: A behavioral and event-related potential study in humans. *Journal of Clinical Neurophysiology, 26*(6), 407–413.
Yang, W., Li, Q., Ochi, T., Yang, J., Gao, Y., Tang, X., et al. (2013). Effects of auditory stimuli in the horizontal plane on audiovisual integration: An event-related potential study. *PloS One, 8*(6), e66402. https://doi.org/10.1371/journal.pone.0066402.
Zeljko, M., & Grove, P. M. (2016). Sensitivity and bias in the resolution of stream-bounce stimuli. *Perception, 46*(2), 1–27.
Zhao, S., Wang, Y., Jia, L., Feng, C., Liao, Y., & Feng, W. (2017). Pre-coincidence brain activity predicts the perceptual outcome of the streaming/bouncing motion display. *Scientific Reports, 7*, 8832. https://doi.org/10.1038/s41598-017-08801-5.
Zhao, S., Wang, Y., Xu, H., Feng, C., & Feng, W. (2018). Early cross-modal interactions underlie the audiovisual bounce-inducing effect. *NeuroImage, 174*, 208–218.
Zhou, F., Wong, V., & Sekuler, R. (2007). Multisensory integration of spatio-temporal segmentation cues: One plus one does not always equal two. *Experimental Brain Research, 180*(4), 641–654.
Zvyagintsev, M., Nikolaev, A. R., Sachs, O., & Mathiak, K. (2011). Early attention modulates perceptual interpretation of multisensory stimuli. *NeuroReport, 22*(12), 586–591.