



# Temporal Reference, Attentional Modulation, and Crossmodal Assimilation

Yingqi Wan and Lihan Chen\*

School of Psychological and Cognitive Sciences and Beijing Key Laboratory of Behavior and Mental Health, Peking University, Beijing, China

Crossmodal assimilation effect refers to the prominent phenomenon by which ensemble mean extracted from a sequence of task-irrelevant distractor events, such as auditory intervals, assimilates/biases the perception (such as visual interval) of the subsequent task-relevant target events in another sensory modality. In current experiments, using visual Ternus display, we examined the roles of temporal reference, materialized as the time information accumulated before the onset of target event, as well as the attentional modulation in crossmodal temporal interaction. Specifically, we examined how the global time interval, the mean auditory inter-intervals and the last interval in the auditory sequence assimilate and bias the subsequent percept of visual Ternus motion (element motion vs. group motion). We demonstrated that both the *ensemble (geometric) mean* and the *last interval* in the auditory sequence contribute to bias the percept of visual motion. Longer mean (or last) interval elicited more reports of group motion, whereas the shorter mean (or last) auditory intervals gave rise to more dominant percept of element motion. Importantly, observers have shown dynamic adaptation to the temporal reference of crossmodal assimilation: when the target visual Ternus stimuli were separated by a long gap interval after the preceding sound sequence, the assimilation effect by ensemble mean was reduced. Our findings suggested that crossmodal assimilation relies on a suitable temporal reference on adaptation level, and revealed a general temporal perceptual grouping principle underlying complex audio-visual interactions in everyday dynamic situations.

**Keywords:** temporal window, temporal ventriloquism effect, central tendency effect, assimilation, attention

## OPEN ACCESS

Daya Shankar Gupta,  
Camden County College,  
United States

Keita Mitani,  
Tokyo Institute of Technology, Japan  
Hakan Karşilar,  
Koç University, Turkey

Lihan Chen  
CLH@pku.edu.cn

20 February 2018  
16 May 2018  
05 June 2018

Wan Y and Chen L (2018) Temporal Reference, Attentional Modulation, and Crossmodal Assimilation. *Front. Comput. Neurosci.* 12:39. doi: 10.3389/fncom.2018.00039

## INTRODUCTION

Multisensory interaction has been traditionally revealed to take place over a narrowed window time—i.e., within a presumed “temporal window” (Meredith et al., 1987; Powers et al., 2009; Vroomen and Keetels, 2010; Wallace and Stevenson, 2014; Gupta and Chen, 2016). For example, paired sound/tactile events presented in temporal proximity to paired visual events can alter the perceived interval between the visual stimuli, and hence bias the perception of visual apparent motion (Keetels and Vroomen, 2008; Chen et al., 2010; Shi et al., 2010). The above illusions have been typically known as temporal ventriloquism (Chen and Vroomen, 2013). Studies on temporal ventriloquism indeed suggested that crossmodal events appearing in temporal proximities have higher probabilities of “correlation” and even “causation” relations (Ernst and Di Luca, 2011; Parise et al., 2012). Based on those relations, sensory events with higher functional priorities (such

as “precision” in timing) would calibrate/attract the counterpart events (with lower functional appropriateness) from the other modalities, give rise to successful *multisensory integration*. During the integration, multisensory events within a presumed short time window will largely obey the “assumption of unity,” in which the coherent representation of multiple events become possible when they have been deemed as coming from a common source (Vatakis and Spence, 2007, 2008; Misceo and Taylor, 2011; Chuen and Schutz, 2016; Chen and Spence, 2017). As a result, the effectiveness of crossmodal interaction is enhanced.

However, the presumed “temporal window” for integration has often been violated in many ecological scenarios. Take an example: upon hearing the whistle of a running car behind us, after a decent long delay, we can know exactly what kind of the “car” is approaching and then make prompt avoidance. This indicates that humans can adaptively use the prior knowledge and employ the temporal/spatial information (including environmental cues associated with the sound) to facilitate the perceptual decision. This daily scenario, however, imposes a great challenge for human perception. How are perceptual grouping and correspondences between events achieved when the crossmodal events are separated both in longer temporal ranges and with larger temporal disparities? Moreover, for the longer temporal range, observers have difficulties in memorizing all the events and the processing of the sensory properties (including time information) would probably exceed their working memory capacities (Cowan, 2001; Klemen et al., 2009; Klemen and Chambers, 2012; Cohen et al., 2016). Therefore, the efficiency of crossmodal interaction will be reduced accordingly. The complex timing scenario as well as the challenge for time cognition also stems from the variance of the multiple time intervals. In short temporal range (such as around 2 s), human observers could discriminate the short temporal intervals when the coefficient of variance (i.e., “CV,” the ratio of the interval deviation to its baseline value) is less than 0.3. The discrimination ability is greatly reduced when the CV is above 0.3 (Allan, 1974; Getty, 1975; Penney et al., 2000).

To cope with the above constraints, human observers adopt one of the efficient perceptual strategies—“ensemble coding” to process the mean properties of multiple events. For example, people can extract the mean rhythm of a given sound sequence and use this information to allocate visual attention and facilitate the detection of target events (Miller et al., 2013). Recent studies have shown that this averaging process is highly dependent on the temporal reference. The temporal reference included the generally global time interval before the onset of target event(s), the variabilities of the multiple intervals and the critical information of the last interval (Jones and McAuley, 2005; Acerbi et al., 2012; Cardinal, 2015; Karaminis et al., 2016). One compelling example is the central tendency effect within the broader framework of Bayesian optimization (Jazayeri and Shadlen, 2010; Shi et al., 2013; Shi and Burr, 2016; Roach et al., 2017), whereby incorporating the mean of the statistical distribution in the estimation would assimilate the estimates toward the mean (Jazayeri and Shadlen, 2010; Burr et al., 2013; Karaminis et al., 2016). For example, the estimation of a target property, such as the duration of an event, is assimilated toward

to the mean duration of previously encountered target events (i.e., event history) (Nakajima et al., 1992; Burr et al., 2013; Shi et al., 2013; Roach et al., 2017). The central tendency effect indicates that human observers exploit predictive coding using the averaged sensory properties (Shi and Burr, 2016). The predictive coding framework states that the brain produces a Bayesian estimate of the environment (Friston, 2010). A strong mismatch between the prediction and the actual sensory input leads to an update of the internal model, and could trigger observable changes in perceptual decision. During this updating, attentional process can be considered as a form of predictive coding to establish an expectation of the moments in time until the task-relevant, to be integrated stimulus inputs arrive (Klemen and Chambers, 2012). On the other hand, the

2010). Here we extended the Ternus temporal ventriloquism paradigm to investigate the temporal crossmodal ensemble coding. We implemented five experiments to address this issue. Experiments 1 and 2 examined the role of temporal window-interval gap between the onset of sound sequence and the onset of target Ternus display, to show the temporal constraints of central tendency effect. Experiment 3 compared the central tendency effect with the recency effect, by manipulating both the mean auditory interval and the last auditory interval. In Experiment 4, we fixed the last interval to be equal to the transitional threshold of perceiving element vs. group motion in the pretest, and manipulated the mean auditory inter-interval to show a genuine central tendency effect during crossmodal assimilation. In Experiment 5, we implemented dual-tasks and asked observers to perform the visual Ternus task while fulfilling a concurrent task of counting oddball sounds. Overall, the current results revealed that crossmodal central tendency effect is subject to the temporal reference (including the length of global time interval, the mean interval and the last interval for a given sound sequence) but less dependent on attentional modulation.

## MATERIALS AND METHODS

### Participants

A total of 60 participants (14, 13, 7, 12, 14 in Experiments 1–5), ages ranging from 18 to 33 years, took part in the main experiments. A *post-hoc* power estimation has shown the statistical powers are generally approaching or above 0.8 for the given sample sizes. All observers had normal or corrected-to-normal vision and reported normal hearing. The experiments were performed in compliance with the institutional guidelines set by the Academic Affairs Committee, School of Psychological and Cognitive Sciences, Peking University. The protocol was approved by the Committee for Protecting Human and Animal Subjects, School of Psychological and Cognitive Sciences, Peking University. All participants gave written informed consent in accordance with the Declaration of Helsinki, and were paid for their time on a basis of 40 CNY/hour, i.e., 6.3 US dollars/hour.

### Apparatus and Stimuli

The experiments were conducted in a dimly lit (luminance: 0.09 cd/m<sup>2</sup>) room. Visual stimuli were presented at the center of a 22-inch CRT monitor (FD 225P) at a screen resolution of 1024 × 768 pixels and a refresh rate of 100 Hz. Viewing distance was 57 cm, maintained by using a chin rest. A Ternus display consisted of two stimulus frames, each containing two black discs (10.30 cd/m<sup>2</sup>; disc diameter and separation between discs: 1.6° and 3° of visual angle, respectively) presented on a gray background (16.3 cd/m<sup>2</sup>). The two frames shared one element location at the center of the monitor, while containing two other elements located at horizontally opposite positions relative to the center (see Figure 1A). Each frame was presented for 30 ms; the inter-stimulus interval (ISI<sub>V</sub>) between the two frames was randomly selected from the range of 50–230 ms, with a step size of 30 ms.

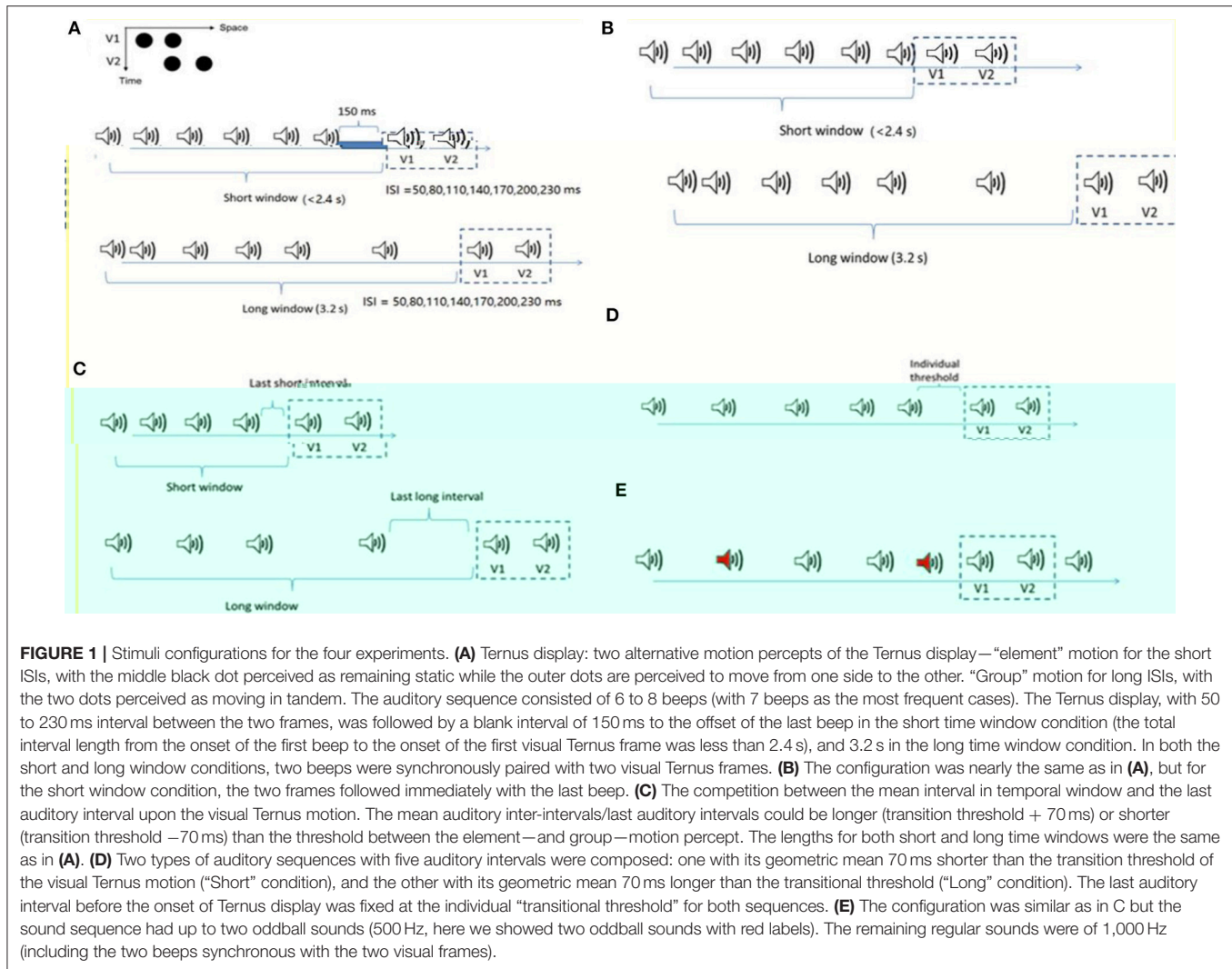
Mono sound beeps (1,000 Hz pure tone, 65 dB SPL, 30 ms, except in Experiment 5 where pure tones with pitches of either 1,000 Hz or 500 Hz were given) were generated and delivered via

an M-Audio card (Delta 1010) to a headset (Philips, SHM1900). No ramps were applied to modulate the shape of the tone envelope. To ensure accurate timing of the auditory and visual stimuli, the duration of the visual stimuli and the synchronization of the auditory and visual stimuli were controlled via the monitor's vertical synchronization pulses. The experimental program was written with Matlab (Mathworks Inc.) and the Psychophysics Toolbox (Brainard, 1997; Kleiner et al., 2007).

## Experimental Design

### Practice

Prior to the formal experiment, participants were familiarized with Ternus displays of either typical “element motion” (with an interval of 50 ms) or “group-0.056(w)0.) ounto nk(y)-426.311(t



auditory sequence consisting a variable number of 6–8 beeps was presented (see below for the details of the onset of Ternus-display frames relative to that of the auditory sequence). A trial began with the presentation of a central fixation marker, randomly for 300 to 500 ms. After a 600-ms blank interval, the auditory train and the visual Ternus frames were presented (see Figure 1A), followed sequentially by a blank screen of 300 to 500 ms and a screen with a question mark at the screen center prompting participants to indicate the type of motion they had perceived: element vs. group motion (non-speeded response). During the experiment, observers were simply asked to indicate the type of visual motion (“element” or “group” motion) that they perceived, while ignoring the beeps. After the response, the next trial started following a random inter-trial interval of 500 to 700 ms.

In Experiment 1, the visual Ternus frames were preceded by an auditory sequence of 6–8 beeps with the geometric mean of inter-stimulus interval [ISIA(uditory), i.e.,  $ISI_A$ ], manipulated to be 70 ms shorter than, or 70 ms longer than the transition threshold estimated in the pre-test. The [ISIV(isual), i.e.,  $ISI_V$ ] between the two visual Ternus frames was randomly selected

from one of the following seven intervals: 50, 80, 110, 140, 170, 200, and 230 ms. The total auditory sequence consisted of 6–8 beeps. Visual Ternus frames were presented on most of all trials (672 trials in total) following the last beep; the remaining were catch trials (72 trials) in which the frames were inset in the sound sequence to break up anticipatory processes. For the short time window of the auditory sequence, the time interval from the onset of the first beep to the onset of the first visual frame was less than 2.4 s, and the gap interval between the onset of the last beep and the onset of the first Ternus frame was 150 ms. For the long time window, the total interval from the onset of the sound to the first visual frame was 3.2 s. In both the short and long window conditions, two beeps were synchronously paired with two visual Ternus frames. All the trials were randomized and organized in 12 blocks (62 trials for each block).

In Experiment 2, the settings were the same as in Experiment 1, except for the condition: the visual frames were following immediately with the onset of the last beep.

In Experiment 3, we introduced two factors of interval modulations: the mean interval of temporal window and the last

auditory interval. The mean auditory inter-intervals and the last auditory intervals could be larger (transition threshold + 70 ms) or shorter (transition threshold - 70 ms) than the threshold between the element- and group- motion percept. Therefore, there were four combinations of the “interval” conditions: both the mean interval and the last interval were shorter (i.e., “MeanSLastS”); the mean interval was shorter but the last interval was longer (“MeanSLastL”); the mean was longer but the last interval was shorter (“MeanLLastS”); and both the mean interval and the last interval were longer (“MeanLLastL”). The onset of the two visual Ternus frames (30 ms) was accompanied by a (30-ms) auditory beep (i.e.,  $ISI_V = ISI_A$ ).

In Experiment 4 we compared two auditory sequences: one with its geometric mean 70 ms shorter than the transition threshold of the visual Ternus motion (hereafter the “Short” condition), and the other with its geometric mean 70 ms longer than the transitional threshold (hereafter the “Long” condition). Instead of randomization of the five auditory intervals (excluding the final synchronous auditory interval with the visual Ternus interval), the last auditory interval before the onset of Ternus display was fixed at the “transitional threshold” for both sequences. The rest four intervals were chosen randomly such that the coefficient of variance (CV) of the auditory sequence was in the range between 0.1 and 0.2, which is the normal range of CV for human observers (Allan, 1974; Getty, 1975; Penney et al., 2000). By this manipulation, we expected to minimize the influence of the potential recency effect caused by the last auditory interval. The audiovisual Ternus frames were appended at the end of these sequences for 85.7% trials (with 672 trials out of 784 trials), in which the Ternus display appeared at the end of the sound sequence (the “onset” of first visual frame was synchronized with 6th beep). The remaining were 112 catch trials, in which 56 trials had the Ternus displays at the beginning of the sound sequence (i.e., the “onset” of the first visual frame was synchronized with the second beep), and the rest 56 trials at middle temporal locations (i.e., the “onset” of the first visual frame was synchronized with the 4th beep). Those catch trials were used to avoid potential anticipatory attending to the visual events appearing at the end of the sound sequence. The total 784 trials were randomized and organized in 14 blocks, with each of 56 trials.

In Experiment 5, we used three types of auditory sequences, in which the mean auditory interval was either *shorter than*, *equal to* or *longer than* the individual transitional threshold of Ternus motion. The auditory sequence consisted of 8 to 10 beeps, including those accompanying the two visual Ternus frames, with the latter being inserted mainly at the 6th–7th positions (504 trials), and followed by 0–2 beeps (number selected at random), to minimize expectations for the onset of the visual Ternus frames. Two of the beeps (the 6th and the 7th) were synchronously paired with two visual Ternus frames which were separated by a visual ISI ( $ISI_V$ ) that varied from 50 to 230 ms (for the critical beeps,  $ISI_V = ISI_A$ ). There were up to two oddball tones (500 Hz) in the sound sequence, while the remaining regular sounds were of 1,000 Hz (including the two beeps synchronous with the two visual frames). Participants completed a dual-task in which they not only made discriminations of the

Ternus display (“element motion” vs. “group motion”) but also reported the number of oddball sounds (0–2) (Figure 1).

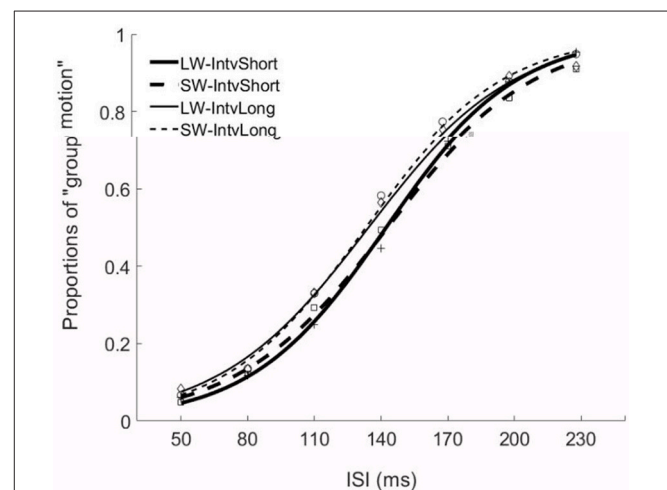
## RESULTS

### Experiment 1: The Effect of Short Temporal Window (With a Temporal Gap Between Auditory Sequence and Visual Ternus) vs. Long Temporal Window

The PSEs for the short window and long window were 149.4 ( $\pm 5.6$ , standard error) ms and 141.2 ( $\pm 4.8$ ) ms. The main effect of temporal window was significant,  $F_{(1, 13)} = 6.878$ ,  $p = 0.021$ ,  $\eta_g^2 = 0.346$ . The PSEs for short interval and long interval were 145.5 ( $\pm 5.2$ ) ms and 145.0 ( $\pm 4.8$ ) ms, the main effect of mean interval was not significant,  $F_{(1, 13)} = 0.120$ ,  $p = 0.735$ ,  $\eta_g^2 = 0.009$ . The interaction effect between factors of window and interval was not significant,  $F_{(1, 13)} = 1.033$ ,  $p = 0.328$ ,  $\eta_g^2 = 0.074$ . For the JNDs, both the main effects of temporal window and mean interval were not significant,  $F_{(1, 13)} = 3.419$ ,  $p = 0.087$ ,  $\eta_g^2 = 0.208$  and  $F_{(1, 13)} = 0.089$ ,  $p = 0.770$ ,  $\eta_g^2 = 0.007$ . And the interaction effect between the two factors was not significant,  $F_{(1, 13)} = 2.863$ ,  $p = 0.114$ ,  $\eta_g^2 = 0.180$  (Figures 2, 4).

### Experiment 2: The Effect of Short Temporal Window (Without a Gap Between Auditory Sequence and Visual Ternus) vs. Long Temporal Window

The PSEs for the short window and long window were 168.7 ( $\pm 6.2$ ) ms and 156.2 ( $\pm 5.7$ ). The PSE for short window was larger



**FIGURE 2** | Psychometric curves for Experiment 1. Mean proportions of group-motion responses were plotted as a function of the probe visual interval (ISI), and fitted psychometric curves, were plotted for the auditory sequences with the different lengths of temporal windows and with different (geometric) mean intervals relative to the individual transition thresholds. SW-IntvLong, Short window with long mean auditory inter-interval; SW-IntvShort, Short window with short mean auditory inter-interval; LW-IntvLong, Long window with long mean auditory inter-interval. LW-IntvShort, long window with short mean auditory inter-interval.

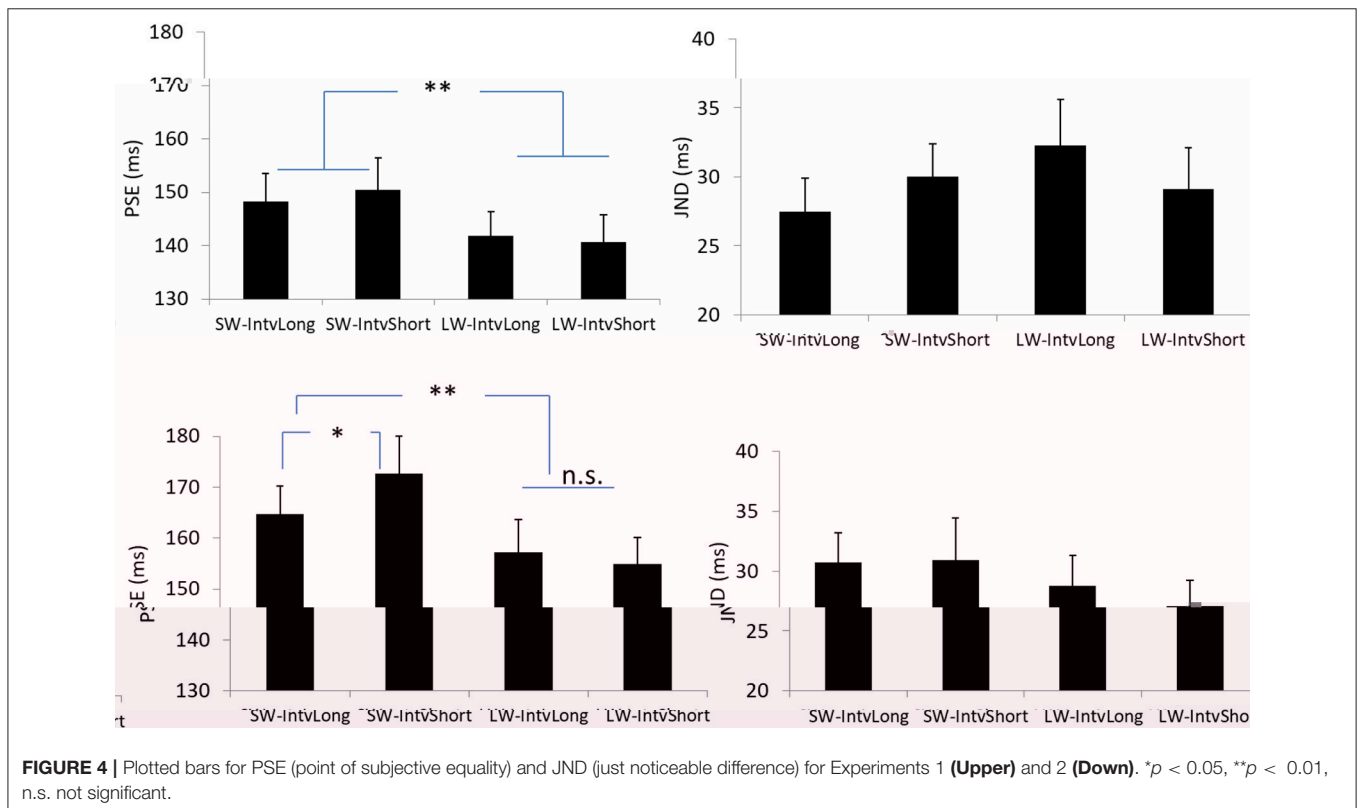
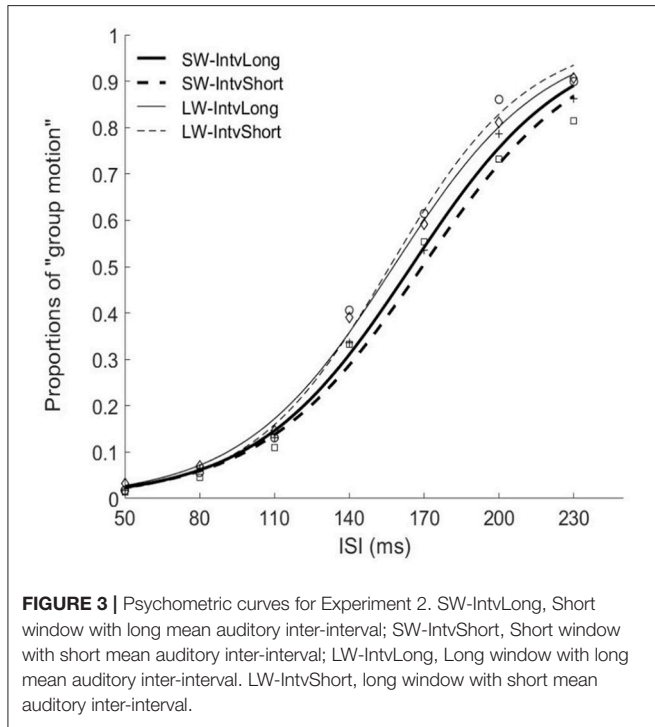
than the one in long window,  $F_{(1, 12)} = 20.860, p = 0.001, \eta_g^2 = 0.635$ . The PSEs for short interval and long interval were  $163.8 (\pm 6.0)$  ms and  $161.0 (\pm 5.8)$ , the main effect of mean interval

was not significant,  $F_{(1, 12)} = 1.869, p = 0.197, \eta_g^2 = 0.135$ . Importantly, the interaction effect between factors of window and interval was significant,  $F_{(1, 12)} = 5.090, p = 0.044, \eta_g^2 = 0.298$ . Further simple effect analyses showed that for short interval, the PSE in short window ( $172.7 \pm 7.3$  ms) was larger than the one ( $154.9 \pm 5.3$  ms) in long window,  $p = 0.001$ . For long interval, the PSE in short window ( $164.7 \pm 5.5$  ms) was larger than the one ( $157.3 \pm 6.4$  ms) in long window,  $p = 0.034$ . On the other hand, for the short window, the PSE in short interval ( $172.7 \pm 7.3$  ms) was larger than the one in long interval ( $164.7 \pm 5.5$  ms),  $p = 0.044$ . However, for the long window, the PSEs are equal in both intervals ( $154.9$  vs.  $157.3$  ms for short and long intervals),  $p = 0.377$ .

For the JNDs, both the main effects of temporal window and mean interval were not significant [ $F_{(1, 12)} = 2.479, p = 0.141, \eta_g^2 = 0.171$  and  $F_{(1, 12)} = 0.282, p = 0.605, \eta_g^2 = 0.023$ ]. The interaction effect between the two factors was not significant,  $F_{(1, 12)} = 0.408, p = 0.535, \eta_g^2 = 0.033$  (Figures 3, 4).

### Experiment 3: Central Tendency Effect vs. Last Interval

The PSEs for the short mean interval and long mean interval were  $143.2 (\pm 7.4)$  ms and  $135.3 (\pm 9.5)$ . The main effect of mean interval was significant,  $F_{(1, 6)} = 9.070, p = 0.024, \eta_g^2 = 0.602$ . The PSEs for short last interval and long last interval were  $155.8 (\pm 9.7)$  ms and  $122.6 (\pm 7.5)$  ms, respectively. The main effect of last interval was significant,  $F_{(1, 6)} = 65.970, p = 0.000, \eta_g^2 = 0.917$ . The interaction effect between factors of mean interval



and last interval was not significant,  $F_{(1, 6)} = 0.195$ ,  $p = 0.674$ ,  $\eta_g^2 = 0.031$ . For the JNDs, the JND in short last interval ( $24.8 \pm 1.3$  ms) was larger than the one in long last interval ( $21.5 \pm 1.6$  ms),  $F_{(1, 6)} = 11.590$ ,  $p = 0.014$ ,  $\eta_g^2 = 0.659$ . However, the main effect of mean interval was not significant,  $F$

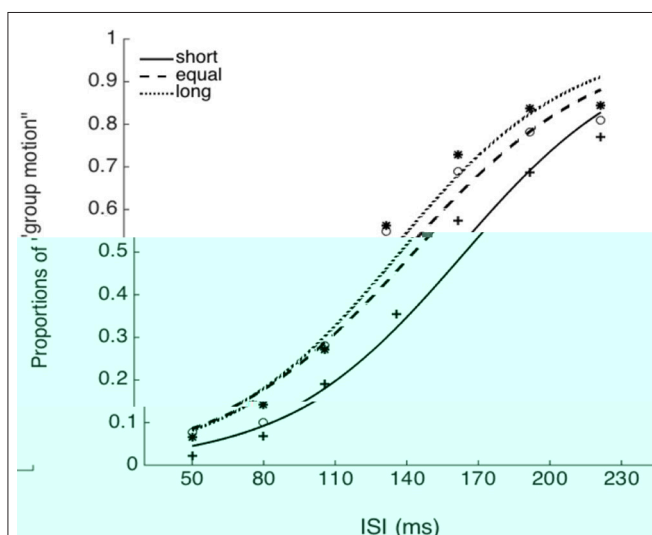
“equal” and “long” conditions,  $p = 0.411$ . The PSEs were equal for both “baseline” and “equal” condition,  $p = 0.603$ , and were equal between “baseline” and “long” conditions,  $p = 1$ .

The JNDs for the *baseline*, *short*, *equal*, and *long* intervals were  $32.2 (\pm 3.7)$ ,  $39.3 (\pm 5.1)$ ,  $44.9 (\pm 7.0)$ , and  $40.0 (\pm 4.4)$  ms, respectively. The main effect of mean interval was not significant,  $F(3, 39) = 2.741$ ,  $p = 0.056$ ,  $\eta_g^2 = 0.174$  (Figures 8, 9).

The mean correct rate for reporting the number of oddball sounds was  $83.0 \pm 3.1\%$ , one sample *T*-test with comparison of 50% showed the correct rate was above the chance level [ $t(13) = 10.518$ ,  $p = 9.984 \times 10^{-8}$ ].

## DISCUSSION

Central tendency, the tendency of judgments of quantitative properties (lengths, durations etc) for given stimuli to gravitate toward their mean, is one of the most robust perceptual effects. The present study has shown that perceptual averaging of temporal property—auditory intervals, assimilates the visual interval between the two Ternus-display frames, and biases the perception of Ternus apparent motion (either to be dominant “element motion” or dominant “group motion”). This finding is consistent with the large body of literature on temporal-context and central tendency effects, within the broader framework of Bayesian optimization (Jazayeri and Shadlen, 2010; Shi et al., 2013; Roach et al., 2017), whereby incorporating the mean of the statistical distribution in the estimation would assimilate the estimates toward the mean—known as “central tendency effect” (Jazayeri and Shadlen, 2010; Burr et al., 2013; Karaminis et al., 2016).



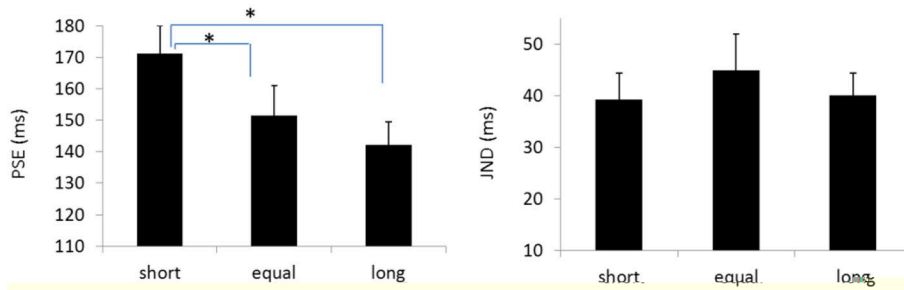
**FIGURE 8** | Psychometric curves for Experiment 5. Short (solid line), the mean auditory inter-interval is shorter than the PSE for visual Ternus motion; Equal (dashed line), the mean auditory inter-interval is equal to the PSE for visual Ternus motion; Long (dotted line), the mean auditory inter-interval is longer than the PSE for visual Ternus motion. The PSE (“transitional threshold”) of Ternus motion was established by a pre-test for each individual.

By using the paradigm of temporal ventriloquism and the probe of visual Ternus display (Chen et al., 2010; Shi et al., 2010; Chen and Vroomen, 2013), we have previously shown that the auditory capture effect upon the visual events, in which the perceived visual interval was biased by concurrently presented auditory events. Observers tended to report the illusory visual (apparent motion) percepts with the concurrent presence of auditory beeps. However, the visual-auditory integration effect is subject to the temporal reference, i.e., the time interval between the critical visual probe and the sound sequence, the mean auditory interval and the critical interval between the last auditory stimulus and the onset of visual events. In our current setting, when the total time interval between the onset of auditory signal and the onset of visual events was above 3 s (3.2 s), it gave rise to a diminished central tendency effect. On the contrary, when this time interval was less than 2.4 s, the shortened time reference increased the likelihood of central tendency effect—materialized in the effect of “geometric” perceptual averaging for auditory intervals upon the visual Ternus motion. These findings indicate a general temporal framework of crossmodal integration. As stated in a theoretical construct of temporal perception, known as the “subjective present”—a mechanism of temporal integration binds successive events into perceptual units of 3 s duration (Pöppel, 1997). Such a temporal integration, which is automatic and pre-semantic, is also operative in movement control and other cognitive activities. In this hierarchical temporal model, the temporal reference for temporal binding could be extended but limited within 3 s, together with a memory store (Pöppel, 1997; Pöppel and Bao, 2014). When the framework exceeds 3 s, the integration of the preceding auditory interval information could be decayed, which hence makes the auditory assimilation effect reduced.

Interestingly, even with the presumed short temporal window (within 2.4 s), by inserting a short temporal gap (150 ms) between the onset of the very last beep and the onset of the first visual frame, we found the central tendency effect was reduced, and the effect was similar to the results in long temporal window condition (3.2 s). This finding suggests that the “imminent” and most recent (“immediate”) temporal gap before the target visual event is critical for the development of the central tendency effect. This inference is further substantiated by the results from Experiments 2 and 3. In Experiment 2, with the configuration of “short window,” we eliminated the short gap (150 ms) between the onset of the last beep and the onset of the visual frames. We found that the central tendency effect (short mean interval. vs. long mean interval) reappeared, though it still remains absent in the condition of “long window.” Moreover, in Experiment 3, we further found that the assimilation effect of the last interval dominates that of the mean auditory interval. This indicates that the last auditory interval wins the competition over the mean interval in driving the crossmodal assimilation.

However, the central tendency effect was less dependent on attentional modulation. Using the dual-tasks of reporting the percept of visual Ternus motion and the number of oddball stimuli [i.e., identifying the number of 500 Hz beep(s) within a sound sequence], we again found the central tendency effect was robust. The observers have invested large attentional resources to





**FIGURE 9** | Plotted bars for PSE (point of subjective equality) and JND (just noticeable difference) for Experiments 5, \* $p < 0.05$ .

obtain the decent performance of counting the oddball sounds. Nevertheless, the performance of crossmodal assimilation effect still survived. Therefore, the central tendency effect as shown in the present study, has demonstrated its automatic and attentional-less demanding nature during crossmodal interaction (Vroomen et al., 2001; Wahn and König, 2015).

The current study has some limitations. Indeed, the temporal reference before the target visual Ternus display includes intervals composed by stimuli with different configurations. The auditory sequence was organized by filled-durations with multiple beeps, and there was a transition of intra-modal perceptual grouping (with sounds) to cross-modal grouping when the last beep was followed by the onset of the first visual Ternus frame (with audiovisual events) (Burr et al., 2013). However, the “critical” time window for multisensory integration was presented as an “empty interval” between the two visual frames. Therefore, the visual probe we adopted in current experimental paradigm might restrict the manifestation of assimilation effect, which was probably due to the differential timing sensitivities to the “filled-duration” in auditory sequence vs. “empty-duration” in the visual probe (Rammsayer and Lima, 1991; Grondin, 1993; Rammsayer, 2010). Moreover, the temporal window, as shown in the auditory sequence, covaried with the mean ISIs (mean auditory intervals). This potential confound remains even although we have manipulated the comparisons of durations between the mean ISIs and the critical interval between the two visual frames (Experiments 1, 2, 3, and 5), and tried to tease apart the “central tendency effect” vs. “recency effect” by fixing the last intervals. Further research is needed to elucidate this point.

Taken together, the current study has shown that crossmodal assimilation in temporal domain is shaped by the temporal reference, in which the observers use the temporal information by dynamically averaging the intervals (as they unfold in time sequence) and exploiting the last interval before the target events. The central tendency effect in temporal domain, similar to the central effect associated with other sensory properties

such as weights and hues, is adaptively subject to the frame of reference (Hollingworth, 1910; Helson, 1947, 1948; Helson

## REFERENCES

- Acerbi, L., Wolpert, D. M., and Vijayakumar, S. (2012). Internal representations of temporal statistics and feedback calibrate motor-sensory interval timing. *PLoS Comput. Biol.* 8:e1002771. doi: 10.1371/journal.pcbi.1002771
- Allan, L. G. (1974). Psychophysical theories of duration discrimination. *Percept. Psychophys.* 16, 26–34. doi: 10.3758/BF03203244
- Alsius, A., Mottonen, R., Sams, M. E., Soto-Faraco, S., and Tiippana, K. (2014). Effect of attentional load on audiovisual speech perception: evidence from ERPs. *Front. Psychol.* 5:727. doi: 10.3389/fpsyg.2014.00727
- Alsius, A., Navarra, J., Campbell, R., and Soto-Faraco, S. (2005). Audiovisual integration of speech falters under high attention demands. *Curr. Biol.* 15, 839–843. doi: 10.1016/j.cub.2005.03.046
- Block, R. A., and Gruber, R. P. (2014). Time perception, attention, and memory: a selective review. *Acta Psychol.* 149, 129–133. doi: 10.1016/j.actpsy.2013.11.003
- Brainard, D. H. (1997). The psychophysics toolbox. *Spat. Vis.* 10, 433–436.

- a “preexperimental” frame of reference. *Percept. Psychophys.* 24, 161–167. doi: 10.3758/BF03199543
- Parise, C. V., Spence, C., and Ernst, M. O. (2012). When correlation implies causation in multisensory integration. *Curr. Biol.* 22, 46–49. doi: 10.1016/j.cub.2011.11.039
- Penney, T. B., Gibbon, J., and Meck, W. H. (2000). Differential effects of auditory and visual signals on clock speed and temporal memory. *J. Exp. Psychol. Hum. Percept. Perform.* 26, 1770–1787. doi: 10.1037/0096-1523.26.6.1770
- Pöppel, E. (1997). A hierarchical model of temporal perception. *Trends Cogn. Sci.* 1, 56–61. doi: 10.1016/S1364-6613(97)01008-5
- Pöppel, E., and Bao, Y. (2014). “Temporal windows as a bridge from objective to subjective time,” in *Subjective Time*, eds D. Lloyd and V. Arstila (Cambridge, MA: MIT Press), 241–261.
- Powers, A. R., Hevey, M. A., and Wallace, M. T. (2012). Neural correlates of multisensory perceptual learning. *J. Neurosci.* 32, 6263–6274. doi: 10.1523/JNEUROSCI.6138-11.2012
- Powers, A. R., Hillock, A. R., and Wallace, M. T. (2009). Perceptual training narrows the temporal window of multisensory binding. *J. Neurosci.* 29, 12265–12274. doi: 10.1523/JNEUROSCI.3501-09.2009
- Rammsayer, T. H. (2010). Differences in duration discrimination of filled and empty auditory intervals as a function of base duration. *Atten. Percept. Psychophys.* 72, 1591–1600. doi: 10.3758/APP.72.6.1591
- Rammsayer, T. H., and Lima, S. D. (1991). Duration discrimination of filled and empty auditory intervals: cognitive and perceptual factors. *Percept. Psychophys.* 50, 565–574. doi: 10.3758/BF03207541
- Roach, N. W., McGraw, P. V., Whitaker, D. J., and Heron, J. (2017). Generalization of prior information for rapid Bayesian time estimation. *Proc. Natl. Acad. Sci. U.S.A.* 114, 412–417. doi: 10.1073/pnas.1610706114
- Sherif, M., Taub, D., and Hovland, C. I. (1958). Assimilation and contrast effects of anchoring stimuli on judgments. *J. Exp. Psychol.* 55, 150–155. doi: 10.1037/h0048784
- Shi, Z., and Burr, D. (2016). Predictive coding of multisensory timing. *Curr. Opin. Behav. Sci.* 8, 200–206. doi: 10.1016/j.cobeha.2016.02.014
- Shi, Z., Chen, L., and Muller, H. J. (2010). Auditory temporal modulation of the visual Ternus effect: the influence of time interval. *Exp. Brain Res.* 203, 723–735. doi: 10.1007/s00221-010-2286-3
- Shi, Z., Church, R. M., and Meck, W. H. (2013). Bayesian optimization of time perception. *Trends Cogn. Sci.* 17, 556–564. doi: 10.1016/j.tics.2013.09.009
- Stevenson, R. A., Siemann, J. K., Schneider, B. C., Eberly, H. E., Woynaroski, T. G., Camarata, S. M., et al. (2014). Multisensory temporal integration in autism spectrum disorders. *J. Neurosci.* 34, 691–697. doi: 10.1523/JNEUROSCI.3615-13.2014
- Stevenson, R. A., Zemtsov, R. K., and Wallace, M. T. (2012). Individual differences in the multisensory temporal binding window predict susceptibility to audiovisual illusions. *J. Exp. Psychol. Hum. Percept. Perform.* 38, 1517–1529. doi: 10.1037/a0027339
- Sugano, Y., Keetels, M., and Vroomen, J. (2010). Adaptation to motor-visual and motor-auditory temporal lags transfer across modalities. *Exp. Brain Res.* 201, 393–399. doi: 10.1007/s00221-009-2047-3
- Sugano, Y., Keetels, M., and Vroomen, J. (2012). The Build-up and transfer of sensorimotor temporal recalibration measured via a synchronization task. *Front. Psychol.* 3:246. doi: 10.3389/fpsyg.2012.00246
- Sugano, Y., Keetels, M., and Vroomen, J. (2016). Auditory dominance in motor-sensory temporal recalibration. *Exp. Brain Res.* 234, 1249–1262. doi: 10.1007/s00221-015-4497-0
- Talsma, D., Doty, T. J., and Woldorff, M. G. (2007). Selective attention and audiovisual integration: is attending to both modalities a prerequisite for early integration? *Cereb. Cortex* 17, 679–690. doi: 10.1093/cercor/bhk016
- Talsma, D., Senkowski, D., Soto-Faraco, S., and Woldorff, M. G. (2010). The multifaceted interplay between attention and multisensory integration. *Trends Cogn. Sci.* 14, 400–410. doi: 10.1016/j.tics.2010.06.008
- Tang, X., Wu, J., and Shen, Y. (2016). The interactions of multisensory integration with endogenous and exogenous attention. *Neurosci. Biobehav. Rev.* 61, 208–224. doi: 10.1016/j.neubiorev.2015.11.002
- Ternus, J. (1926). Experimentelle untersuchungen über phänomenale identität. *Psychol. Forsch.* 7, 81–136. doi: 10.1007/BF02424350
- Thomas, D. R., and Jones, C. G. (1962). Stimulus generalization as a function of the frame of reference. *J. Exp. Psychol.* 64, 77–80. doi: 10.1037/h0043304
- Thomas, D. R., and Strub, H. (1974). Adaptation level and the central tendency effect in stimulus generalization. *J. Exp. Psychol.* 103, 466–474. doi: 10.1037/h0037184
- Thomas, D. R., Svinicki, M. D., and Vogt, J. (1973). Adaptation level as a factor in human discrimination learning and stimulus generalization. *J. Exp. Psychol.* 97, 210–219. doi: 10.1037/h0033908
- Treutwein, B., and Strasburger, H. (1999). Fitting the psychometric function. *Percept. Psychophys.* 61, 87–106. doi: 10.3758/BF03211951
- Vatakis, A., and Spence, C. (2007). Crossmodal binding: evaluating the “unity assumption” using audiovisual speech stimuli. *Percept. Psychophys.* 69, 744–756. doi: 10.3758/BF03193776
- Vatakis, A., and Spence, C. (2008). Evaluating the influence of the ‘unity assumption’ on the temporal perception of realistic audiovisual stimuli. *Acta Psychol.* 127, 12–23. doi: 10.1016/j.actpsy.2006.12.002
- Vroomen, J., Bertelson, P., and de Gelder, B. (2001). The ventriloquist effect does not depend on the direction of automatic visual attention. *Percept. Psychophys.* 63, 651–659. doi: 10.3758/BF03194427
- Vroomen, J., and Keetels, M. (2010). Perception of intersensory synchrony: a tutorial review. *Atten. Percept. Psychophys.* 72, 871–884. doi: 10.3758/APP.72.4.871
- Wahn, B., and König, P. (2015). Audition and vision share spatial attentional resources, yet attentional load does not disrupt audiovisual integration. *Front. Psychol.* 6:1084. doi: 10.3389/fpsyg.2015.01084
- Wallace, M. T., and Stevenson, R. A. (2014). The construct of the multisensory temporal binding window and its dysregulation in developmental disabilities. *Neuropsychologia* 64, 105–123. doi: 10.1016/j.neuropsychologia.2014.08.005
- Wichmann, F. A., and Hill, N. J. (2001). The psychometric function: I. fitting, sampling, and goodness of fit. *Percept. Psychophys.* 63, 1293–1313. doi: 10.3758/BF03194544

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Wan and Chen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.