

ORIGINAL ARTICLE

Rapid audiovisual temporal recalibration across children and adults

Yifan Zhao^{1,2†} , Runzhou Wang^{1,2†}  and Hong-Yan Bi^{1,2} 

¹State Key Laboratory of Cognitive Science and Mental Health, Institute of Psychology, Chinese Academy of Sciences, Beijing 100101, China and ²Department of Psychology, University of Chinese Academy of Sciences, Beijing 100049, China

Corresponding author: Hong-Yan Bi; Email: bihy@psych.ac.cn

(Received 13 April 2024; revised 6 January 2025; accepted 26 February 2025)

Abstract

Audiovisual temporal integration ability, reflected by the size of the temporal binding window (TBW), plays an important role in reading. The audiovisual TBW is not fixed, but dynamically changes during the integration process, this is referred to as rapid temporal recalibration. To investigate the rapid audiovisual temporal recalibration ability across age and its correlation with reading, the present study conducted simultaneity judgment (the index includes Δ PSS and Δ TBW) tasks involving speech (Experiment 1; children: $M_{\text{age}} = 10.70$, adults: $M_{\text{age}} = 24.52$) and non-speech (Experiment 2; children: $M_{\text{age}} = 10.19$, adults: $M_{\text{age}} = 24.26$) audiovisual stimuli in native Mandarin-Chinese-speaking child and adult groups ($n = 36$ in each group). Results showed that children's Δ PSS and Δ TBW for speech stimuli were comparable to those of adults. However, when examining trial-by-trial changes in TBW during the integration process, a gap between children and adults was evident. Besides, for non-speech stimuli, children significantly differed from adults in both Δ PSS indicators and the integration process. Moreover, for both children and adults, the correlation and regression analysis showed that the rapid audiovisual temporal recalibration ability of both speech and non-speech stimuli explained reading fluency uniquely after controlling TBW, age, and gender.

Keywords: Audiovisual temporal integration; developmental status; dynamic temporal course; rapid audiovisual temporal recalibration; reading ability

Introduction

In everyday activities such as conversation, humans receive inputs from auditory (sound) and visual (lip movements) channels, and integrate them into a coherent

[†]These authors have contributed equally to this work.

This work was supported by the STI 2030—Major Projects (2021ZD0200500), the National Natural Science Foundation of China (32371119), and the 2023 Key Project of the “14th Five-Year Plan” Scientific Research Program of the National Language Commission (ZDI145-83).

perceptual representation; the aforementioned process is commonly referred to as audiovisual integration (Alsus et al., 2005; Koelewijn et al., 2010). Audiovisual temporal integration refers to the cognitive process by which individuals perceptually integrate asynchronous audiovisual stimuli into a coherent and meaningful unified event (Stevenson & Wallace, 2013; Stevenson et al., 2012). Temporal factors (such as synchrony, duration, speed, and rhythmic structure) play pivotal roles in effectively integrating visual and auditory signals (Fujisaki et al., 2004; van Wassenhove et al., 2007). However, although temporal proximity is deemed crucial in audiovisual integration (van der Burg et al., 2008), the synchronous presence of audiovisual stimuli is not necessary for this process (Lewkowicz, 1996). When a pair of audiovisual stimuli is presented within the temporal binding window (TBW), they undergo integration (Dixon & Spitz, 1980; Vroomen & Keetels, 2010). Humans are typically able to perceive audiovisual stimuli with a few hundred milliseconds of asynchrony as being presented synchronously (Beneventi et al., 2010; Birch & Belmont, 1964; van Wassenhove et al., 2007). On the contrary, the audiovisual stimuli that are well separated in time and fall outside the TBW are often perceived as irrelevant independent information (Zhou et al., 2020).

The audiovisual temporal integration ability can be gauged by the magnitude of TBW (e.g., Francisco et al., 2014; Hairston et al., 2005), or by indicators that reflect the magnitude of TBW, such as the mean proportion of synchronous responses (e.g., Francisco, Groen, et al., 2017; Francisco, Jesse, et al., 2017) and the threshold of time perception (e.g., just noticeable differences, JND; Mossbridge et al., 2017). TBW refers to a temporal interval, during which two asynchronous stimuli will be integrated and perceived as one event (Dixon & Spitz, 1980; Vroomen & Keetels, 2010). An abnormally narrowed TBW impairs the integration of related stimuli; conversely, an abnormally widened TBW hinders individuals from distinguishing unrelated asynchronous stimuli and mistakenly integrating them into one event (Francisco et al., 2014; Francisco, Jesse, et al., 2017; Hairston et al., 2005; Virsu et al., 2003). However, only focusing on participants' average-level audiovisual temporal integration processing overlooks its dynamic changes over time. Navarra et al. (2005, 2007, 2009) have found that the audiovisual TBW is not fixed but undergoes dynamic changes throughout the integration process, including three stages: temporal window expansion, the point of subjective simultaneity (PSS) shift, and temporal window restoration, known as temporal recalibration. PSS refers to a stimulus onset asynchrony (SOA) in which an individual is most likely to subjectively perceive the stimuli as synchrony (Navarra et al., 2005, 2007, 2009). The expansion and contraction of the temporal window depend on the previous sensory experience (Powers et al., 2009). Specifically, the audiovisual temporal processing will be dynamically adjusted based on prior sensory experience; in an asynchronous audiovisual environment, the PSS will shift toward the asynchronous direction (Fujisaki et al., 2004). For example, after watching a movie in which the auditory and visual elements are not synchronous for a while, the SOA in which an individual is most likely to perceive the stimuli as synchrony (i.e., PSS) will align with the SOA between visual and auditory stimuli in the movie, thus no longer feeling uncomfortable. When the temporal recalibration can be completed in a short time (e.g., 1 trial), it is called rapid temporal recalibration (van der Burg et al., 2013).

This short-term plasticity (the ability to rapidly adapt to changing sensory statistics of environments) plays an important role in audiovisual temporal integration (Noel et al., 2017). De Nier et al. (2017) and van der Burg et al. (2015) have found that the TBW will significantly change (initially increase and subsequently decrease) with the progress of experiments (the increase in the number of trials). The change of TBW during the process of audiovisual temporal integration has also been substantiated by numerous training studies (e.g., Powers et al., 2009, 2012; Theves et al., 2020), thereby reflecting its plasticity. It is evident that a comprehensive assessment of the change process of TBW can provide more informative insights compared to simply aggregating all trials to obtain an overall measure of TBW. Several empirical studies have indicated that despite certain individuals possessing a typical multi-sensory integration ability at the average level, the process of integration (e.g., rapid audiovisual temporal recalibration) may be different from typically developing individuals (Noel et al., 2017; Zaidel et al., 2015). Therefore, the rapid audiovisual temporal recalibration may serve as a more sensitive indicator, considering the more process information provided by this recalibration (e.g., Noel et al., 2017). Investigating the rapid audiovisual temporal recalibration holds great importance in further elucidating the underlying mechanism of audiovisual temporal integration.

Reading is the cognitive process of integrating visual and auditory input information (Pammer & Vidyasagar, 2005). The phonetic information and graphemic information of written language are integrated into the form of graphemes–phonemes binding, and this binding is automated through repeated memory (Chein & Schneider, 2005; Dijkstra et al., 1989), and finally fluent reading is acquired. Gori et al. (2020) emphasized the significance of an appropriate audiovisual TBW in facilitating the accurate binding of pronunciation to print when learning to read. Therefore, it can be inferred that audiovisual temporal integration is a crucial influencing factor in reading. This view has also been confirmed by many empirical studies. Francisco, Groen, et al. (2017) found that the proportion of participants' synchronous response to speech and non-speech audiovisual asynchronous stimuli was positively correlated with the frequency of reading errors, and could effectively explain the variance in reading errors. Later, both behavioral and EEG studies conducted by Mossbridge et al. (2017) supported that audiovisual temporal integration ability was positively correlated with reading comprehension performance and could explain 16% and 25% of reading comprehension variance respectively. Mossbridge and colleagues suggested that audiovisual temporal integration processing might contribute to synchronizing phonological representations with visual orthographic processing in the process of orthographic-to-phonetic decoding (Mossbridge et al., 2017). Liu et al. (2019) did similar research on the Chinese writing system and found that the audiovisual cross-modal temporal order judgment task performance not only directly impacted reading ability, but also indirectly impacted it through rapid naming and orthographic processing (Liu et al., 2019). Therefore, previous results substantiated the correlation between audiovisual temporal integration and reading. The rapid audiovisual temporal recalibration capability represents a dynamic adjustment process in response to environmental changes and a measure of variability, which ultimately impacts stable temporal representations, such as the TBW (Noel et al.,

2016). In addition, there exists a positive correlation between individuals' temporal integration ability and rapid audiovisual temporal recalibration ability (Harvey *et al.*, 2014; Noel *et al.*, 2016; van der Burg *et al.*, 2013). Specifically, rapid recalibration may serve as the underlying mechanism facilitating temporal integration by permitting swift adjustments of TBW and PSS in response to successive environmental stimuli for successful integration (Noel *et al.*, 2017). Consequently, there may be a significant correlation between rapid audiovisual temporal recalibration and reading. A strong rapid audiovisual temporal recalibration ability may enhance the process of reading acquisition by dynamically modulating temporal windows and PSS. Specifically, during real-world reading acquisition, individuals inevitably encounter instances of asynchrony between sound and image (e.g., when learning via video on an iPad). At such times, individuals with skillful temporal recalibration abilities can flexibly adjust their TBW and PSS in response to constant environmental changes, thereby acquiring accurate morphology-pronunciation correspondence rules and ultimately achieving optimal reading ability. However, there remains a lack of empirical evidence, especially in Chinese reading. This study sought to clarify this question.

Compared with children, there is a diminished correlation observed between audiovisual temporal integration ability and reading in adults (children: Wu, 2020; adults: Francisco, Groen, *et al.*, 2017). Therefore, the relationship between rapid audiovisual temporal recalibration and reading may also be modulated by age group. From a developmental perspective, audiovisual temporal integration matures earlier for non-speech stimuli (flashes and beeps) than for speech stimuli (speech) (Noel *et al.*, 2016). Since the temporal integration ability and rapid audiovisual temporal recalibration ability of speech and non-speech audiovisual stimuli are correlated (Harvey *et al.*, 2014; Noel *et al.*, 2016; van der Burg *et al.*, 2013), the rapid audiovisual temporal recalibration may have a similar development trend with audiovisual temporal integration. Several studies have found that the temporal integration ability for speech and non-speech audiovisual stimuli in childhood has already comparable to adults (Chen *et al.*, 2016; Hillock-Dunn *et al.*, 2016; Lewkowicz & Flom, 2014). Additionally, the rapid audiovisual temporal recalibration ability matures earlier than the audiovisual temporal integration ability (Burr & Gori, 2012; Noel *et al.*, 2016). Therefore, the rapid audiovisual temporal recalibration ability may be comparable to adults in childhood. But so far, in the context of Chinese culture, only Zhou *et al.* (2020) have studied the effects of age and stimulus type on rapid audiovisual temporal recalibration. They found that adults exhibited significant rapid audiovisual temporal recalibration for both speech and non-speech stimuli. However, adolescents only exhibited significant temporal recalibration for speech stimuli, while failing to exhibit this significant effect for non-speech stimuli. It can be observed that within the context of Chinese culture, individual's rapid audiovisual temporal recalibration for speech stimuli matures earlier than that for non-speech stimuli. However, Zhou *et al.* (2020) only examined the differences between adolescents and adults. To date, there has been a lack of exploration into the difference in rapid audiovisual temporal recalibration ability between children and adults, including its impact from stimulus types (speech and non-speech) and whether its relationship with reading is modulated by age group.

In summary, investigating the cross-age difference of rapid audiovisual temporal recalibration and its relationship with reading is crucial for elucidating the temporal course and mechanisms underlying audiovisual temporal integration, as well as understanding how audiovisual temporal integration explains the variance of reading abilities. The current body of research is insufficient to establish whether the rapid audiovisual temporal recalibration observed in childhood is comparable to that observed in adults, and its relationship with reading remains unconfirmed. Therefore, the present study focused on the ability of rapid audiovisual temporal recalibration and made a further exploration of three aspects: age, stimulus types, and their correlation with reading. In Experiment 1, both children and adults underwent the audiovisual simultaneous judgment (SJ) task with speech stimuli; in Experiment 2, both children and adults underwent the audiovisual SJ task with non-speech stimuli. Built upon previous efforts, the present study hypothesizes that: (1) children's rapid audiovisual temporal recalibration ability matures earlier for speech stimuli than for non-speech stimuli; (2) rapid audiovisual temporal recalibration ability is related to reading ability.

Experiment 1

Materials and methods

Participants

Upon ethical approval and informed consent, we recruited 36 children and 36 adults from China. The children (18 females) were recruited from primary schools, with a mean age of 10.70 years old ($SD = 1.88$). The adults (27 females) were recruited online from colleges, with a mean age of 24.52 years old ($SD = 1.82$). All participants were native Chinese speakers with Asians as their race who possessed normal hearing and either normal vision or corrected-to-normal vision. None of the participants had a history of neurological disease or psychiatric disorders.

Audiovisual SJ task with speech stimuli

Participants performed a standard SJ task with multisensory stimuli (judge the synchrony of audiovisual stimuli; e.g., Francisco et al., 2014; Francisco, Groen, et al., 2017; Francisco, Jesse, et al., 2017). The audiovisual clips featuring a female speaker uttering the single syllable “ba” were used (i.e., speech stimuli, see Figure 1). The visual stimuli (1200 × 680 pixels; 30 frames per second) had a duration ranging from 2090 to 2370 ms. Each video clip contained the entire process of articulation, including pre-articulatory gestures. The auditory stimuli had a sampling rate of 44.1 kHz and 16-bit depth, with a duration of approximately 225 ms. The set of SOAs utilized were 0, ±40, ±80, ±120, ±160, ±200, ±240, and ±280 ms, for the auditory and visual stimuli. The negative values indicate that the presentation of auditory stimulus precedes that of visual stimulus, while positive values indicate that the presentation of visual stimulus precedes that of auditory stimulus. The SJ task comprised a total of 150 trials, with 10 trials for each SOA condition. The inter-trial interval between consecutive trials was set at 500 ms. We obtained written consent from the female speaker to utilize these videos in experiments and publications.

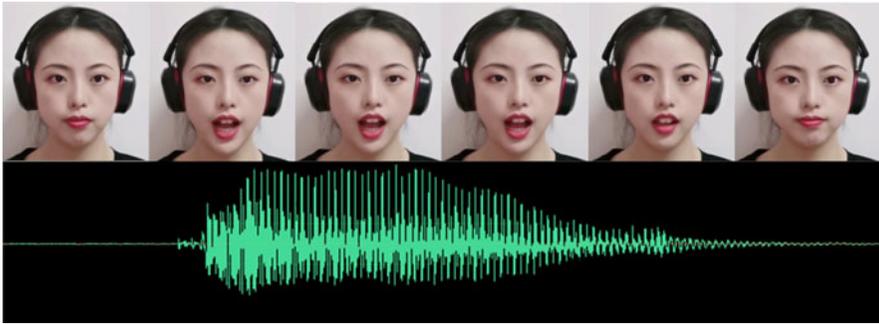


Figure 1. The Speech Stimulus in Experiment 1.

Note. Top to bottom represents the individual frames from dynamic visual stimuli and the auditory waveform for the stimulus utilized (i.e., SOA = 0 ms).

Reading measures

Participants' reading fluency was assessed using the reading test used by Qian *et al.* (2015) and Meng *et al.* (2019, 2022), which has been widely used in studies of children and adults. The test demonstrated high levels of reliability and validity, with an internal consistency coefficient (Cronbach's alpha) ranging from 0.93 to 0.97 (Meng *et al.*, 2019, 2022). A printed list containing 160 high-frequency single characters (with a mean frequency of 233.71 times per million) was provided to participants, who were instructed to read aloud as many characters as possible within one minute. The total number of characters correctly named was recorded.

Design and general procedure

In the audiovisual SJ task with speech stimuli, a single-factor design was applied, utilizing group (child group and adult group) as the between-subjects factor.

The presentation of stimuli was controlled by the E-prime 3.0 software (E-Prime Psychology Software Tools Inc., Pittsburgh, USA), and the audiovisual clips were presented on a Lenovo Legion 7 15.6-inch liquid crystal display (LCD) screen with a resolution of 1920 × 1080 pixels, 144 Hz refresh rate, and black background. The screen was positioned at a distance of approximately 70 cm from the participants. The auditory stimuli were delivered via headphones. The intensity of the auditory stimuli was individually adjusted for each participant, corresponding to a comfortable hearing level of approximately 73 dB. As shown in Figure 2, each trial commenced with the presentation of a central fixation point, with duration randomized from 500 to 1000 ms. Ten repetitions per SOA condition were presented randomly for the audiovisual clips.

Participants were instructed to indicate whether the audiovisual stimuli were presented synchronously or not by pressing the corresponding response key on the keyboard (i.e., the "f" key for synchronous judgment and the "j" key for asynchronous judgment). Before the formal experiment, eight trials were administered for practice, consisting of four trials with synchronous audiovisual stimuli (SOA = 0 ms), two trials with the largest positive SOA (+280 ms), and two trials with the largest negative SOA (−280 ms). Only after participants reached an accuracy rate of $\geq 75\%$ could they proceed to the formal experiment. It typically

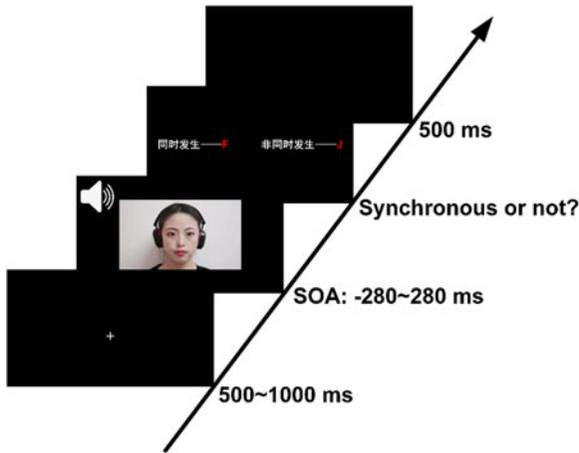


Figure 2. The Procedure of Audiovisual SJ Task in Experiment 1.

took 10–12 min to complete the audiovisual SJ task. Following the audiovisual SJ task, participants completed the reading fluency test to assess their reading ability.

Data analysis

Audiovisual TBW. The proportion of “simultaneity” responses for each SOA condition was used to calculate the magnitude of each participant’s audiovisual TBW in the audiovisual SJ task. The distributions of “synchrony” reports across SOAs were fitted with a Gaussian distribution.

$$P(\text{response}|\text{SOA}) = \text{amp} \times \exp^{-\left(\frac{(\text{SOA}-\text{PSS})^2}{2\text{SD}^2}\right)} \quad (1)$$

In function (1), the mean of the best-fitting distribution was taken as an indicator for PSS, while the standard deviation was taken as an indicator for TBW (Harvey et al., 2014; Noel et al., 2016; Zhou et al., 2020). The data that failed to fit were excluded, and the shape of Gaussian distribution could accurately depict the reports of synchrony (mean adjusted $R^2 > 0.8$, see Table S1; see the data distribution pattern of a few representative participants in Figure S1). It should be noted that there were no participants excluded from the analysis due to failing to fit with a Gaussian distribution. The PSS and TBW were then compared between children and adults using an independent sample t -test. In cases where the assumption of the homogeneity of variances was violated, Welch corrections were applied to adjust the degrees of freedom.

Rapid audiovisual temporal recalibration. To reveal the temporal course of audiovisual temporal integration, the present study first examined the rapid audiovisual temporal recalibration. Separate inter-trial analyses were conducted for the audiovisual SJ task to explore whether the modality order (auditory-leading or visual-leading) in the prior trial (trial_{t-1}) affected the distribution of “synchrony” responses (i.e., whether a rapid audiovisual temporal recalibration occurred) in the current trial (trial_t), as reflected by ΔPSS and ΔTBW which are represent different

manifestations of the same underlying mechanism (van der Burg *et al.*, 2013). Gaussian distributions were fitted to both types of trial_{*t*-1} order [i.e., auditory-leading (trial_{*t*-1}AV) or visual-leading (trial_{*t*-1}VA)] data for each participant to estimate ΔPSS [function (2)] and ΔTBW [function (3)] (e.g., De Nier *et al.*, 2017; Harvey *et al.*, 2014; Noel *et al.*, 2016; Zhou *et al.*, 2020). If rapid audiovisual temporal recalibration occurred, the ΔPSS and/or ΔTBW were expected to be significantly larger than 0. This study conducted a one-sample *t*-test to examine the presence of rapid audiovisual temporal recalibration in children and adults. Subsequently, an independent sample *t*-test was conducted to compare the ΔPSS and ΔTBW between children and adults. In cases where the assumption of the homogeneity of variances was violated, Welch corrections were applied to adjust the degrees of freedom.

$$\Delta\text{PSS} = \text{PSS}_{\text{trial}_{t-1}\text{VA}} - \text{PSS}_{\text{trial}_{t-1}\text{AV}} \quad (2)$$

$$\Delta\text{TBW} = \text{TBW}_{\text{trial}_{t-1}\text{AV}} - \text{TBW}_{\text{trial}_{t-1}\text{VA}} \quad (3)$$

Second, the temporal course analysis was conducted on a total of 150 trials. A sliding-window approach was used to fit Gaussian distribution trial by trial (i.e., trial₁ ~ trial₅₀, trial₂ ~ trial₅₁, . . . , trial₁₀₁ ~ trial₁₅₀) to obtain TBWs that changed with trial (e.g., De Nier *et al.*, 2017; van der Burg *et al.*, 2015), resulting in a total of 101 time points. After excluding data that failed to fit (less than 6% of the total for each condition in each group, as shown in Table S2), the temporal course of audiovisual temporal integration was analyzed via a one-sample *t*-test versus the initial estimate of the given parameter (i.e., TBW on trial₁ ~ trial₅₀). Given the inherent multiple comparisons problem in utilizing a sliding-window approach, the present study corrected for false positives by considering an effect significant by setting $\alpha < .01$ for at least 10 consecutive time points (e.g., De Nier *et al.*, 2017). Referring to van der Burg *et al.* (2015), to fully compare the performance of children and adults, the present study subsequently divided the temporal course into some bins and conducted linear fitting, followed by a comparison of the differences (i.e., analysis of variance, ANOVA) in the slopes of TBWs in each bin between children and adults. The division of bins is exploratory and post-hoc, based on the dynamic trend of TBW (van der Burg *et al.*, 2015). This approach aimed to investigate the differences in the temporal course of audiovisual temporal integration between children and adults. The Greenhouse-Geisser correction was reported whenever Mauchly's test of sphericity was significant. Furthermore, the linear trend at point method was used to substitute any missing values generated during the fitting process (Vercoulen *et al.*, 1994).

The relationship between rapid audiovisual temporal recalibration and reading. The partial correlation analysis was conducted to reveal the correlation between audiovisual temporal integration (TBW), rapid audiovisual temporal recalibration (ΔPSS and ΔTBW), and reading ability while controlling for age and gender. Additionally, hierarchical regressions were conducted to examine the unique contributions of ΔPSS in explaining variance in reading ability after controlling for age, gender, and TBW.

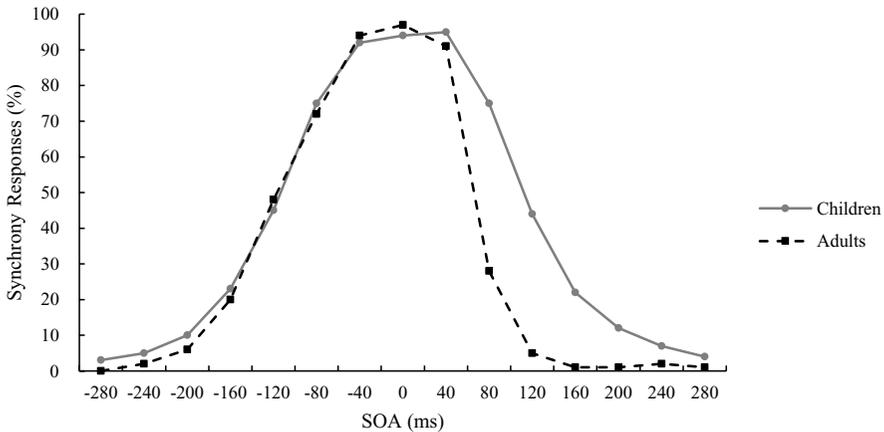


Figure 3. Synchrony Responses by SOAs for Audiovisual Speech Stimuli in Children and Adults Instruction.

Note. Experiment 1 included a sample of 36 children and 36 adults.

Results

The average level of audiovisual temporal integration for speech stimuli in child and adult group

Table 1 and Figure 3 show the mean proportion of responses judged as synchronous responses for both children and adults across each SOA for speech stimuli, suggesting a tendency that children perceive audiovisual speech information as synchrony over a wider temporal window compared to adults. The independent sample *t*-test revealed a significant effect of the group on the width of TBW. The child group ($M = 191.41$, $SD = 41.89$) showed a wider TBW than the adult group ($M = 144.96$, $SD = 23.73$), $t(70) = 5.79$, $p < .001$, Cohen's $d = 1.36$. These results indicated that the average level of audiovisual temporal integration for speech stimuli in children was weaker than that of adults. Moreover, it is noteworthy that the PSS in the adult group exhibited a negative value, indicating a preference for auditory information when perceiving audiovisual asynchronous stimuli. One sample *t*-tests (test value = 0) indicated that the adult group [$M = -26.69$, $SD = 14.65$, $t(35) = -10.93$, $p < .001$, Cohen's $d = 1.82$] demonstrated the auditory processing advantages compared to the child group [$M = 0.30$, $SD = 22.59$, $t(35) = 0.08$, $p = .937$, Cohen's $d = 0.01$].

The rapid audiovisual temporal recalibration of audiovisual temporal integration for speech stimuli in child and adult group

For rapid audiovisual temporal recalibration, the independent sample *t*-test on Δ PSS and Δ TBW revealed no significant effect of group, indicating that there is no discernible difference in the rapid audiovisual temporal recalibration between children and adults [Δ PSS: $t(55.95) = -0.32$, $p = .752$, Cohen's $d = 0.07$; Δ TBW: $t(59.93) = 0.29$, $p = .774$, Cohen's $d = 0.07$]. Furthermore, one sample *t*-tests (test value = 0) indicated the significant positive Δ PSS in both child group [$M = 11.32$, $SD = 21.51$, $t(35) = 3.16$, $p = .003$, Cohen's $d = 0.53$] and adult group

Table 1. Synchrony responses by SOAs and groups for audiovisual speech stimuli

Condition	Children ^a		Adults ^b		<i>t</i>	<i>p</i>	Cohen's <i>d</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>			
SOA = -280 ms	.03	.06	0	0	<i>t</i> (35) = 2.71	.025^c	0.71
SOA = -240 ms	.05	.10	.02	.04	<i>t</i> (45.38) = 1.76	.142	0.39
SOA = -200 ms	.10	.13	.06	.07	<i>t</i> (70) = 1.59	.176	0.38
SOA = -160 ms	.23	.20	.20	.18	<i>t</i> (70) = 0.68	.564	0.16
SOA = -120 ms	.45	.29	.48	.26	<i>t</i> (70) = -0.47	.643	0.11
SOA = -80 ms	.75	.20	.72	.21	<i>t</i> (70) = 0.64	.564	0.15
SOA = -40 ms	.92	.13	.94	.11	<i>t</i> (70) = -0.92	.454	0.17
SOA = 0 ms	.94	.08	.97	.06	<i>t</i> (64.50) = -1.43	.215	0.42
SOA = 40 ms	.95	.08	.91	.11	<i>t</i> (70) = 1.76	.142	0.42
SOA = 80 ms	.75	.19	.28	.18	<i>t</i> (70) = 10.46	<.001	2.54
SOA = 120 ms	.44	.23	.05	.09	<i>t</i> (44.93) = 9.32	<.001	2.23
SOA = 160 ms	.22	.20	.01	.05	<i>t</i> (39.91) = 6.14	<.001	1.44
SOA = 200 ms	.12	.13	.01	.05	<i>t</i> (46.15) = 4.26	<.001	1.12
SOA = 240 ms	.07	.10	.02	.06	<i>t</i> (57.26) = 2.33	.049	0.61
SOA = 280 ms	.04	.07	.01	.02	<i>t</i> (42.88) = 2.76	.025	0.58
<i>M</i> _{All conditions}	.40	.08	.31	.06	<i>t</i> (70) = 5.75	<.001	1.27

Note: When the assumption of the homogeneity of variances was violated, Welch corrections were applied to adjust the degrees of freedom. To correct for multiple comparisons, this study conducted false discovery rate (FDR) correction on the resulting *p* values (same as below; e.g., Benjamini & Hochberg, 1995).

^aChildren: *n* = 36.

^bAdults: *n* = 36.

^cThe bold ones are significant results.

[*M* = 12.63, *SD* = 12.40, *t*(35) = 6.11, *p* < .001, Cohen's *d* = 1.02]. The significant positive Δ PSS indicates that the PSS has adjusted in accordance with the previous trial, demonstrating the presence of rapid audiovisual temporal recalibration (e.g., van der Burg & Goodbourn, 2015; van der Burg et al., 2013, 2015).

As for the other side of rapid audiovisual temporal recalibration, that is, the changes of TBW at trial-by-trial level during the integration process, there was an observed visual trend that the TBW increased first and then remained stable over time in the child group (see Figure 4). The significant differences were detected in width between TBW₁ (i.e., trial₁ ~ trial₅₀, *M* = 167.07, *SE* = 6.95) and the interval including TBW₂₅ (i.e., trial₂₅ ~ trial₇₄) to TBW₁₀₁ (i.e., trial₁₀₁ ~ trial₁₅₀; one sample *t*-test, all *ps* < 0.01). On the contrary, as shown in Figure 4, there was an observed visual trend (rather than a statistical trend) that the TBW increased first and then decreased over time in the adult group. The significant differences were detected in width between TBW₁ (*M* = 132.64, *SD* = 5.72) and the interval including TBW₁₈ (i.e., trial₁₈ ~ trial₆₇) to TBW₃₆ (i.e., trial₃₆ ~ trial₈₅; one sample *t*-test, all *ps* < 0.01).

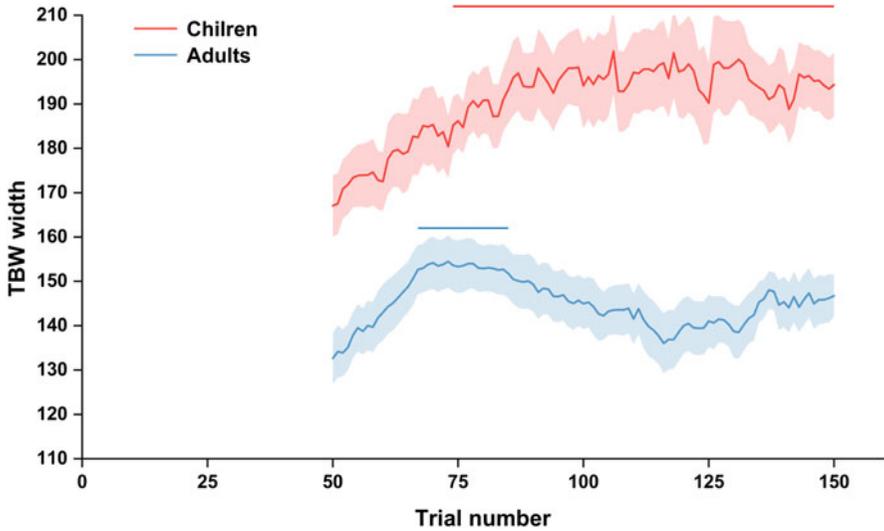


Figure 4. The Temporal Course of Audiovisual Temporal Integration of the TBW for Speech Stimuli in Child and Adult Group Instruction.

Note. The shaded region illustrated the standard error of the mean (SEM) at each trial across the temporal course analysis. To correct for multiple comparisons, this study considered an effect significant at $\alpha < .01$ for at least 10 consecutive trials. Solid bars shown above the temporal course were indicative of at least 10 consecutive trials at which the TBW significantly differed from TBW1 ($\alpha < .01$ for all trials).

To more clearly illustrate the different change process of TBW between children and adults, the present study divided the temporal course into 3 bins (according to Figure 4), performed linear fitting, and subsequently compared the differences in the slopes of TBWs in each bin [bin1: TBW₁ ~ TBW₂₀ (i.e., trial₁ ~ trial₅₀ and trial₂₀ ~ trial₆₉); bin2: TBW₂₁ ~ TBW₆₀ (i.e., trial₂₁ ~ trial₇₀ and trial₆₀ ~ trial₁₀₉); bin3: TBW₆₁ ~ TBW₁₀₁ (i.e., trial₆₁ ~ trial₁₁₀ and trial₁₀₁ ~ trial₁₅₀)] between children and adults. The 2 (group: child group, adult group) \times 3 (bin: bin1, bin2, bin3) ANOVA conducted on the slopes of TBWs revealed a significant main effect of bin [$F(2, 140) = 10.83, p < .001, \eta_p^2 = 0.13$]. Specifically, the slopes of TBWs for bin1 ($M = 0.88, SE = 0.19$) were significantly greater than that for bin2 ($M = -0.07, SE = 0.11$; mean difference = 0.95, $p < .001$) and bin3 ($M = 0.11, SE = 0.13$; mean difference = 0.77, $p = .005$), while there was no difference between the slopes of TBWs for bin2 and bin3 (mean difference = $-0.18, p = .689$). The main effect for the group was not significant [$F(1, 70) = 0.04, p = .835, \eta_p^2 < 0.01$], and there was no significant interaction effect observed [$F(2, 140) = 2.99, p = .054, \eta_p^2 = 0.04$]. These results suggested that, similar to adults, children can rapidly adjust their response to the temporal relationship of the current audiovisual stimuli based on the temporal relationship of the audiovisual stimuli in the prior trial, with a dynamical adjustment of TBW during the audiovisual temporal integration for speech stimuli.

In summary, children around 10 years old exhibit robust rapid audiovisual temporal recalibration for speech stimuli. Although no significant differences are observed between age groups in Δ TBW and Δ PSS, differences in the recalibration

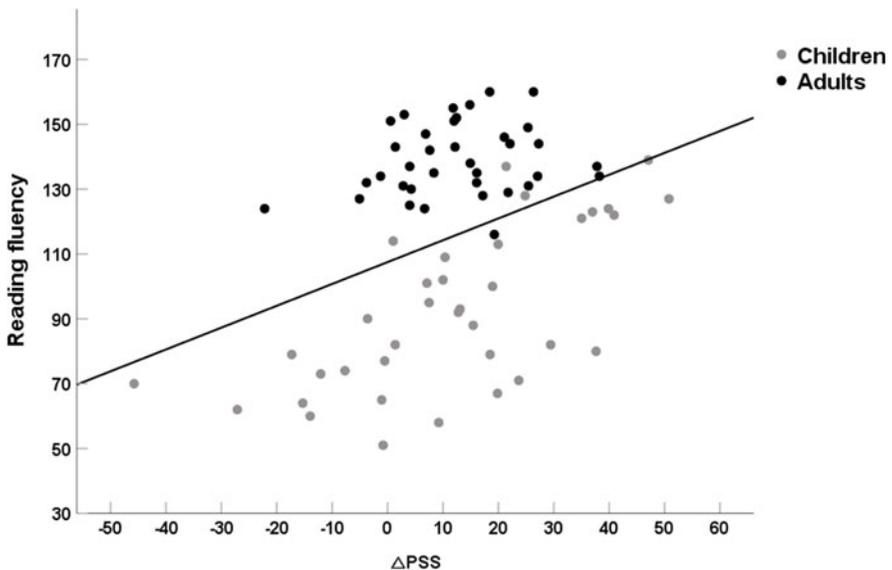


Figure 5. Correlation of Δ PSS and reading fluency in Experiment 1.

process at the trial-by-trial level do exist between the two groups (as shown in Figure 4). Therefore, children's rapid audiovisual temporal recalibration has the potential for further improvement.

The relationship between the temporal course of audiovisual temporal integration for speech stimuli and reading

The relationships between TBW, Δ TBW, Δ PSS, and reading ability were analyzed using partial correlation analysis, with age and gender controlled. There was a marginally significant correlation between TBW and reading fluency, $r = -.19$, $p = .114$. Δ PSS was significantly correlated with reading fluency, $r = .51$, $p < .001$ (see Figure 5). However, the correlation between Δ TBW and reading fluency was not significant, $r = .02$, $p = .859$.

Hierarchical regression analyses were conducted to investigate the impact of the average level of audiovisual temporal integration (with TBW as the indicator) and the temporal course of audiovisual temporal integration (with Δ PSS as the indicator) for speech stimuli on reading ability. Age and gender were controlled, which entered into the regression models as Step 1. Subsequently, TBW entered into the regression models as Step 2. Finally, Δ PSS entered into the regression models as Step 3. The results of the hierarchical regression analyses indicated that after controlling age and gender, Δ PSS explained 9.3% of the variance in reading fluency. When further controlling for TBW, Δ PSS remained a significant predictor of reading fluency, explaining 10.6% of the variance in reading fluency (see more details in Table 2).

Table 2. Hierarchical regression of TBW and Δ PSS on reading fluency in Experiment 1

Variables	ΔR^2	β	<i>t</i>
Age and Gender controlled			
Step 1	.65***		
Age		.76	11.93***
Gender		.05	0.72
Step 2	.09***		
Δ PSS		.31	4.94***
Age, Gender, and TBW controlled			
Step1	.65***		
Age		.67	9.46***
Gender		.02	0.36
Step2	.01		
TBW		-.19	-2.72**
Step3	.11***		
Δ PSS		.33	5.52***

Note: † $p < .1$, * $p < .05$, ** $p < .01$, *** $p < .001$.

Experiment 2

Materials and methods

Participants

Upon ethical approval and informed consent, we recruited 36 children and 36 adults from China. The children (23 females) were recruited from primary schools, with a mean age of 10.19 years old ($SD = 1.94$). The adults (24 females) were recruited online from colleges, with a mean age of 24.26 years old ($SD = 1.98$). All participants were native Chinese speakers with Asians as their race who possessed normal hearing and either normal vision or corrected-to-normal vision. None of the participants had a history of neurological disease or psychiatric disorders. The fact that there was an overlap (30 adults) between the participants in Experiment 2 and Experiment 1 should be acknowledged. To ensure that the task quality was not affected by fatigue and careless attitude, adequate rest time was provided to the adults between the two experiments. Furthermore, a counterbalanced order was implemented for conducting the two experiments.

Audiovisual SJ task with non-speech stimuli

Participants performed a standard SJ task with non-speech audiovisual stimuli (i.e., the dynamic handheld tools, see Figure 6). The visual stimuli (1200×680 pixels; 30 frames per second) consisted of a complete cycle of motion of a hand utilizing a hammer. The videos had a duration ranging from 2090 to 2370 ms. The

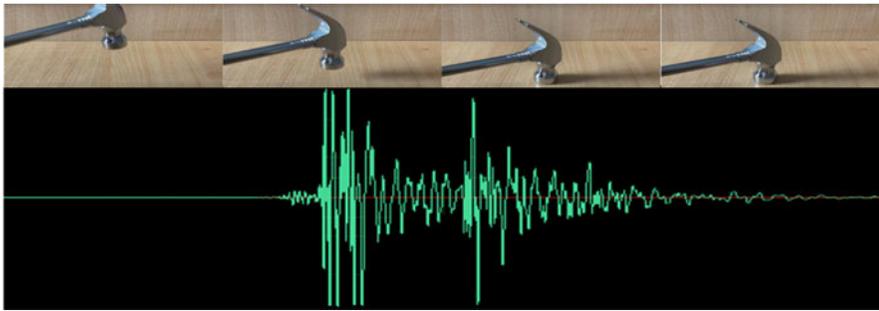


Figure 6. The non-speech stimulus in experiment 2.

Note. Top to bottom represents the individual frames from dynamic visual stimuli and the auditory waveform for the stimulus utilized (i.e., SOA = 0 ms).

auditory stimuli were a congruent hammering noise presented by a sampling rate of 44.1 kHz and 16-bit depth, with a duration of approximately 206 ms. The set of SOA utilized was consistent with Experiment 1. The SJ task comprised a total of 150 trials, with 10 trials for each SOA condition. The inter-trial interval between consecutive trials was set at 500 ms.

Reading measures

The same reading measure was used in Experiment 2 as in Experiment 1.

Design and general procedure

In the audiovisual SJ task with non-speech stimuli, a single-factor design was applied, utilizing group (child group and adult group) as the between-subjects factor.

The presentation of stimuli was controlled by the E-prime 3.0 software (E-Prime Psychology Software Tools Inc., Pittsburgh, USA), and the audiovisual clips were presented on a Lenovo Legion 7 15.6-inch LCD screen with a resolution of 1920 × 1080 pixels, 144 Hz refresh rate, and black background. The distance between the screen and the participants was approximately 70 cm. The auditory stimuli were delivered via headphones. The intensity of the auditory stimuli was individually adjusted for each participant, corresponding to a comfortable hearing level of approximately 73 dB. As shown in Figure 7, each trial commenced with the presentation of a central fixation point, with duration randomized from 500 to 1000 ms. Ten repetitions per SOA condition were presented randomly for the dynamic handheld tools.

A similar audiovisual SJ task (i.e., judge whether the audiovisual stimuli were presented synchronously or not) in Experiment 1 was used in Experiment 2 for non-speech stimuli. The setup of SOA conditions and the setup of practice and formal experiments were the same as in Experiment 1. It typically took 10–12 min to complete the audiovisual SJ task. Following the audiovisual SJ task, participants completed the reading fluency test.

Table 3. Synchrony responses by SOAs and groups for audiovisual non-speech stimuli

Condition	Children ^a		Adults ^b		<i>t</i>	<i>p</i>	Cohen's <i>d</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>			
SOA = -280 ms	.01	.02	.02	.05	<i>t</i> (49.07) = -1.20	.238	0.26
SOA = -240 ms	.02	.05	.01	.02	<i>t</i> (50.28) = 1.86	.069	0.26
SOA = -200 ms	.06	.11	.01	.03	<i>t</i> (41.13) = 2.83	.007^c	0.62
SOA = -160 ms	.10	.15	.03	.07	<i>t</i> (47.87) = 2.37	.022	0.60
SOA = -120 ms	.14	.21	.05	.08	<i>t</i> (46.01) = 2.50	.016	0.57
SOA = -80 ms	.28	.22	.22	.22	<i>t</i> (70) = 1.11	.269	0.27
SOA = -40 ms	.71	.24	.74	.22	<i>t</i> (70) = -0.61	.544	0.13
SOA = 0 ms	.99	.03	.95	.08	<i>t</i> (46.54) = 2.99	.004	0.66
SOA = 40 ms	.97	.06	.98	.05	<i>t</i> (70) = -0.84	.406	0.18
SOA = 80 ms	.88	.13	.79	.13	<i>t</i> (70) = 3.03	.003	0.69
SOA = 120 ms	.50	.32	.30	.21	<i>t</i> (59.71) = 3.18	.002	0.74
SOA = 160 ms	.28	.26	.12	.13	<i>t</i> (51.37) = 3.44	.001	0.78
SOA = 200 ms	.18	.21	.04	.07	<i>t</i> (41.94) = 3.62	.001	0.89
SOA = 240 ms	.10	.12	.02	.05	<i>t</i> (47.33) = 3.49	.001	0.87
SOA = 280 ms	.08	.14	.05	.09	<i>t</i> (70) = 0.98	.332	0.25
<i>M</i> _{All conditions}	.35	.09	.29	.05	<i>t</i> (54.47) = 3.62	.001	0.82

Note: The Welch corrections and multiple comparison corrections were consistent with those in Table 1.

^aChildren: *n* = 36.

^bAdults: *n* = 36.

^cThe bold ones are significant results.

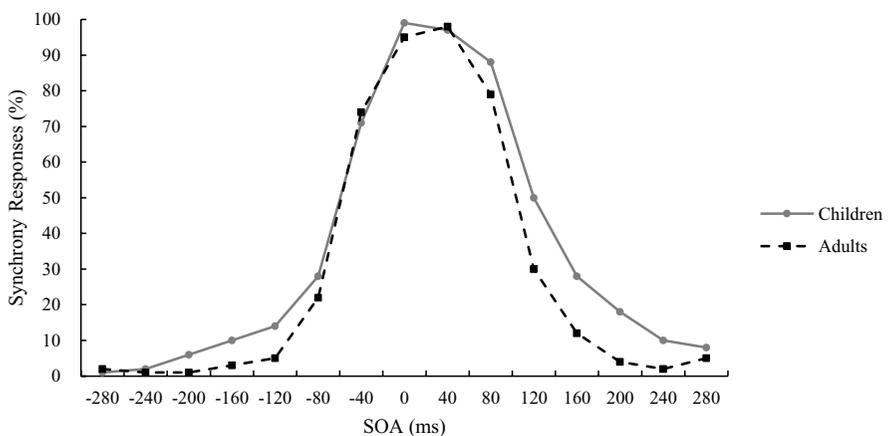


Figure 8. Synchrony responses by SOAs for audiovisual non-speech stimuli in children and adults. Note. Experiment 2 included a sample of 36 children and 36 adults.

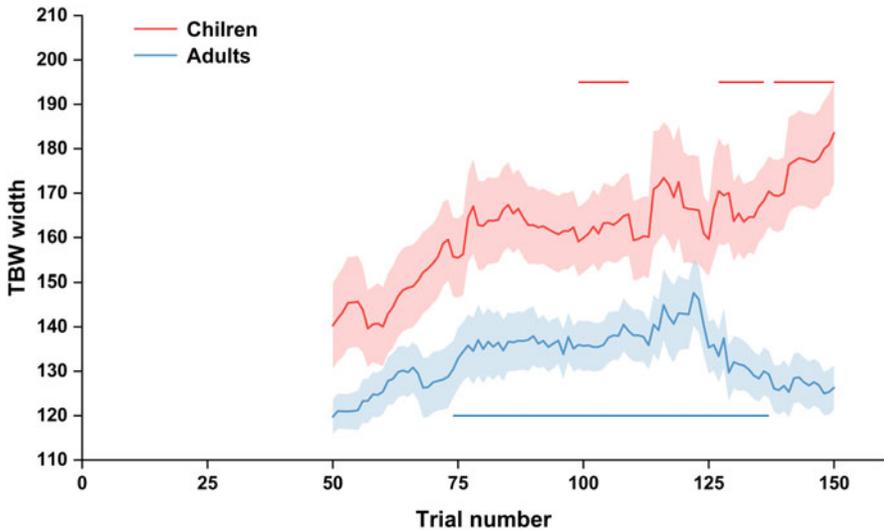


Figure 9. The temporal course of audiovisual temporal integration of the TBW for non-speech stimuli in child and adult groups. *Note.* The shaded region illustrated the standard error of the mean (SEM) at each trial across the temporal course analysis. To correct for multiple comparisons, this study considered an effect significant at $\alpha < .01$ for at least 10 consecutive trials. Solid bars shown above the temporal course were indicative of at least 10 consecutive trials at which the TBW significantly differed from TBW_1 ($\alpha < .01$ for all trials).

auditory processing, as compared to the child group, $t(44.62) = 1.90$, $p = .065$, Cohen's $d = 0.45$.

The temporal course of audiovisual temporal integration for non-speech stimuli in child and adult group

For rapid audiovisual temporal recalibration, the independent sample t -test on Δ PSS revealed a significant effect of group. Adults ($M = 10.83$, $SD = 12.61$) had stronger rapid audiovisual temporal recalibration than children ($M = 4.26$, $SD = 11.88$), $t(70) = 2.28$, $p = .026$, Cohen's $d = 0.54$. However, there was no difference in Δ TBW between children ($M = 0.11$, $SD = 32.79$) and adults ($M = 5.58$, $SD = 31.15$), $t(70) = -0.73$, $p = .470$, Cohen's $d = 0.17$. Furthermore, one sample t -test (test value = 0) indicated the presence of audiovisual temporal recalibration in both child group [Δ PSS: $t(35) = 2.15$, $p = .038$, Cohen's $d = 0.36$] and adult group [Δ PSS: $t(35) = 5.15$, $p < .001$, Cohen's $d = 0.86$].

As for the other side of rapid audiovisual temporal recalibration, that is, the changes of TBW in trial-by-trial level during the integration process, there was an observed visual trend that the TBW first increased, followed by a period of stability, and subsequently increased in the child group (see Figure 9). The differences in width between TBW_1 ($M = 140.30$, $SE = 9.66$) and the interval including TBW_{50} (i.e., trial₅₀ ~ trial₉₉) to TBW_{60} (i.e., trial₆₀ ~ trial₁₀₉), TBW_{78} (i.e., trial₇₈ ~ trial₁₂₇) to TBW_{87} (i.e., trial₈₇ ~ trial₁₃₆), and TBW_{89} (i.e., trial₈₉ ~ trial₁₃₈) to TBW_{101}

(i.e., $\text{trial}_{101} \sim \text{trial}_{150}$; one sample *t*-test, all $ps < 0.01$) were significant. On the contrary, there was an observed visual trend that the TBW initially increased and subsequently decreased over time in the adult group (see more details in Figure 9). The differences in width between TBW_1 ($M = 119.72$, $SD = 3.89$) and the interval including TBW_{25} (i.e., $\text{trial}_{25} \sim \text{trial}_{74}$) to TBW_{88} (i.e., $\text{trial}_{88} \sim \text{trial}_{137}$; one sample *t*-test, all $ps < 0.01$) were significant.

To more clearly illustrate the different change process of TBW between children and adults, the present study divided the temporal course into 3 bins (according to Figure 9), performed linear fitting, and subsequently compared the differences in the slopes of TBWs in each bin [bin1: $\text{TBW}_1 \sim \text{TBW}_{30}$ (i.e., $\text{trial}_1 \sim \text{trial}_{50}$ and $\text{trial}_{30} \sim \text{trial}_{79}$); bin2: $\text{TBW}_{31} \sim \text{TBW}_{70}$ (i.e., $\text{trial}_{31} \sim \text{trial}_{80}$ and $\text{trial}_{70} \sim \text{trial}_{119}$); bin3: $\text{TBW}_{71} \sim \text{TBW}_{101}$ (i.e., $\text{trial}_{71} \sim \text{trial}_{120}$ and $\text{trial}_{101} \sim \text{trial}_{150}$)] between children and adults. The 2 (group: child group, adult group) \times 3 (bin: bin1, bin2, bin3) ANOVA conducted on the slopes of TBWs revealed a significant main effect of bin [$F(1.84, 128.48) = 4.22$, $p = .020$, $\eta_p^2 = 0.06$] and a significant main effect of group [$F(1, 70) = 18.80$, $p < .001$, $\eta_p^2 = 0.21$]. Moreover, there was a significant interaction effect of group and bin, $F(1.84, 128.48) = 3.86$, $p = .027$, $\eta_p^2 = 0.05$. Specifically, simple effect analysis with Sidak correction for multiple comparisons showed that the slopes of TBWs for adults ($M = -0.63$, $SE = 0.19$) were smaller than that for children ($M = 0.64$, $SE = 0.19$) in bin3 (mean difference = -1.27 , $p < .001$), while there were no differences between children (bin1: $M = 0.83$, $SE = 0.25$; bin2: $M = 0.18$, $SE = 0.19$) and adults (bin1: $M = 0.48$, $SE = 0.25$; bin2: $M = 0.15$, $SE = 0.19$) in bin1 (mean difference = -0.35 , $p = .315$) and bin2 (mean difference = -0.02 , $p = .928$). Besides, there were no significant differences between bin1, bin2, and bin3 in the child group; all $ps > .100$. However, within the adult group, the slopes of TBWs in bin1 were significantly greater than that in bin3 (mean difference = 1.10 , $p = .009$), while the slopes of TBWs in bin2 were also significantly greater than that in bin3 (mean difference = 0.78 , $p = .018$). No difference was found between the slopes of TBWs in bin1 and bin2 (mean difference = 0.32 , $p = .723$). Furthermore, one sample *t*-test (test value = 0) indicated no significant slopes of TBWs for bin2 [$t(35) = 1.01$, $p = .319$, Cohen's $d = 0.17$], but did show significant positive slopes of TBWs for bin1 [$t(35) = 3.44$, $p = .002$, Cohen's $d = 0.58$] and bin3 [$t(35) = 3.50$, $p = .001$, Cohen's $d = 0.59$] in child group. However, the adult group exhibited marginally significant positive slopes of TBWs for bin1 [$t(35) = 1.91$, $p = .065$, Cohen's $d = 0.32$], significant negative slopes of TBWs for bin3 [$t(35) = -3.30$, $p = .002$, Cohen's $d = 0.54$], but no significant slopes of TBWs for bin2 [$t(35) = 0.73$, $p = .469$, Cohen's $d = 0.12$]. These results suggested that, compared with children, adults can adjust TBW more flexibly during the audiovisual temporal integration for non-speech audiovisual stimuli, and had a stronger ability to rapidly adjust their response to the temporal relationship of the current audiovisual stimuli based on the temporal relationship in the prior trial.

In summary, these findings suggested that children around the age of 10 already exhibit a strong rapid audiovisual temporal recalibration for non-speech stimuli; however, the significant disparity in this ability between children and adults persists.

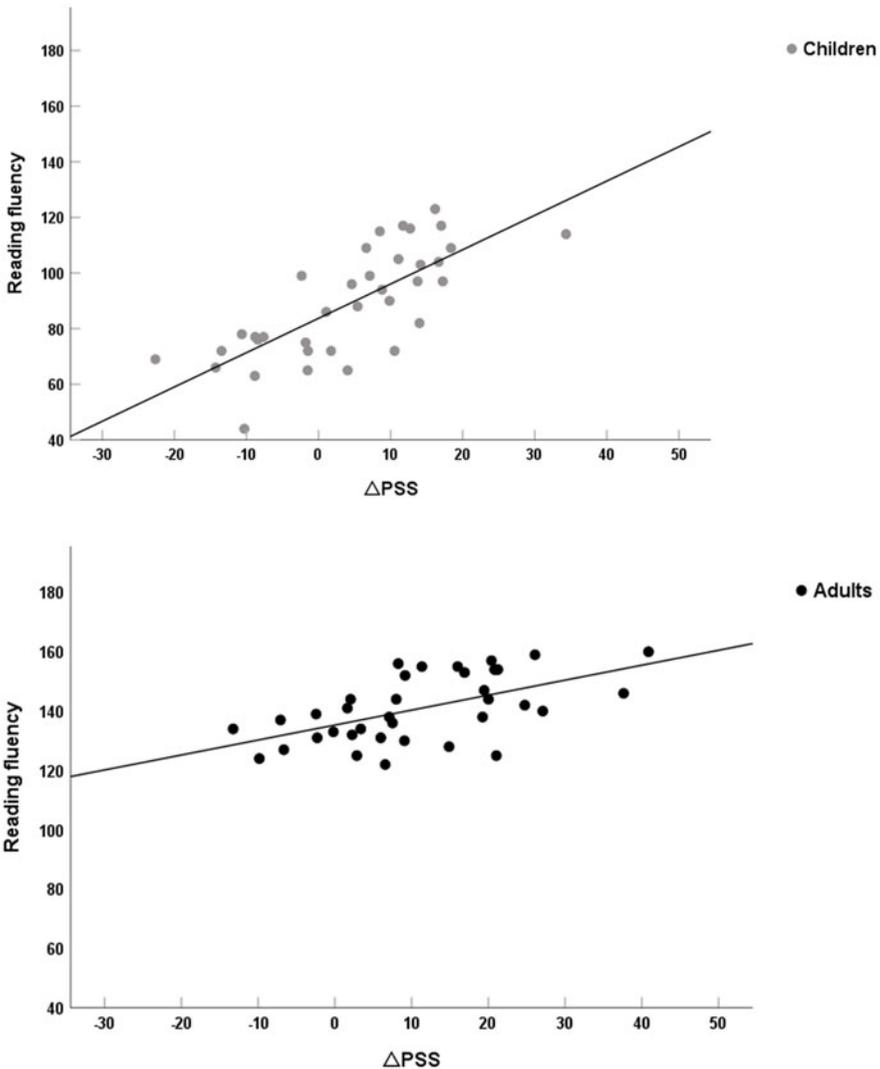


Figure 10. Correlation of Δ PSS and reading fluency in Experiment 2.

The relationship between the temporal course of audiovisual temporal integration for non-speech stimuli and reading

Considering the significant differences in rapid audiovisual temporal recalibration ability between children and adults, the data of children and adults were split, and the partial correlation analysis between TBW, Δ TBW, Δ PSS, and reading fluency was conducted separately. The partial correlation analysis controlling for age and gender showed that children's TBW was significantly correlated with reading fluency, $r = -.37$, $p = .034$, while adults' TBW was marginally significant with reading fluency, $r = -.33$, $p = .058$. In addition, there was a significant positive correlation between children's Δ PSS and reading fluency, $r = .76$, $p < .001$, and

Table 4. Hierarchical regression of TBW and Δ PSS on reading fluency in both children and adults in Experiment 2

Variables	ΔR^2		β		<i>t</i>	
	Children	Adults	Children	Adults	Children	Adults
Age and Gender controlled						
Step 1	<.01	.09				
Age			.15	.05	1.28	0.38
Gender			.00	.25	-0.03	1.78 [†]
Step 2	.58***	.29***				
Δ PSS			.78	.55	6.66***	3.90***
Age, Gender, and TBW controlled						
Step1	<.01	.09				
Age			.11	-.13	1.10	-0.89
Gender			-.03	.17	-0.27	1.26
Step2	.13*	.10 [†]				
TBW			-.36	-.40	-3.61**	-2.83**
Step3	.57***	.32***				
Δ PSS			.78	.57	7.76***	4.51***

Note: [†] $p < .1$, * $p < .05$, ** $p < .01$, *** $p < .001$.

between adults' Δ PSS and reading fluency, $r = .57, p < .001$ (see Figure 10). However, the correlation between Δ TBW and reading fluency was not significant in both children ($r = .09, p = .610$) and adults ($r = -.19, p = .286$).

To explore whether the temporal course of audiovisual temporal integration could explain variations in reading ability, separate hierarchical regression analyses were conducted by group (the dependent variable is reading fluency, and independent variables included the TBW and Δ PSS). The order of independent variables entry into the regression models remained consistent with that in Experiment 1 (i.e., Step 1: age and gender; Step 2: TBW; Step 3: Δ PSS). The results of the hierarchical regression analyses indicated that after controlling age and gender, Δ PSS explained 58.0% and 29.3% of the variance in reading fluency among children and adults respectively. When further controlling for TBW, Δ PSS remained significant for reading fluency, explaining 57.1% and 32.2% of the variance in reading fluency among children and adults respectively (see more details in Table 4; see other regression results in Table S5).

Discussion

The present study compared the rapid audiovisual temporal recalibration using SJ tasks with speech and non-speech stimuli in a sample comprising children and adults. The results showed no significant age-related differences in both Δ TBW and

Δ PSS for speech stimuli. By contrast, significant differences were observed between children and adults in Δ PSS for non-speech stimuli. During the trial-by-trial integration process, there were differences between children and adults for both speech and non-speech stimuli. Besides, the rapid audiovisual temporal recalibration (with Δ PSS as the indicator) for both speech and non-speech stimuli was significantly correlated with reading fluency. Moreover, Δ PSS accounted for the unique variance in reading fluency for both children and adults.

Cross-age changes in rapid audiovisual temporal recalibration abilities

Comparing the results of the two experiments, the present study found that rapid temporal recalibration matures earlier for speech stimuli than for non-speech stimuli. The results align with those of Zhou et al. (2020) in Chinese culture and further extend this line of work to developing children. Specifically, Zhou et al. (2020) found that both adolescents and adults exhibited rapid temporal recalibration to speech stimuli, while only adults showed rapid temporal recalibration to simple non-speech stimuli. However, the development of rapid temporal recalibration abilities in the present study contradicted the expectation that statistical learning of temporal regularities for simple stimuli should develop before more complex stimuli. One possible explanation is that this developmental pattern and earlier maturation for speech-related temporal recalibration may reflect the ethological significance and primacy of language and communicative signals over other simple stimuli (e.g., Murray et al., 2016; Zhou et al., 2020), thereby reaching maturity earlier. Previous studies have found that the TBW of speech stimuli narrowed significantly during the preschool years (Lewkowicz & Flom, 2014) and reached maturity by age 7 (Hillock-Dunn et al., 2016), providing further evidence for the early development and swift maturation of temporal integration processing in speech stimuli. Furthermore, the results of the present study are inconsistent with those of Noel et al. (2016) in Western cultures, as they found that rapid audiovisual temporal recalibration does not reach maturity until adulthood; specifically, the rapid temporal recalibration matured at age 18 for flash beeps and age 29 for syllables. The inconsistent results may stem from cultural differences, as Zhou et al. (2020) highlighted. The study of Noel et al. (2016) was conducted in Western cultures, while the present study and the study of Zhou et al. (2020) investigated age-related differences in the rapid audiovisual temporal recalibration in the Chinese context. The earlier maturation observed in non-western populations may reflect the necessity of considering the potential effect of different cultures on this basic multisensory ability (e.g., Zhou et al., 2020). Interestingly, the present study also found that adults exhibited a preference for referring to auditory information when perceiving both speech and non-speech audiovisual asynchronous stimuli, as compared to children. The previous cross-cultural study suggested that East Asians, compared with Western people, may exhibit a greater reliance on auditory cues during multisensory speech integration (Tanaka et al., 2010). Tanaka et al. (2010) propose that the preference of East Asians for auditory information in speech stimuli may be partly attributed to their limited range of facial muscle movements, which restricts the availability of facial information.

The different processing preferences for visual and auditory inputs in multisensory integration may underlie the differential processing of rapid audiovisual temporal recalibration across Eastern and Western cultures. Besides, it is worth noting that non-speech visual stimuli lack the offset cue (i.e., “hammer lift”), which may impede an individual’s judgment of the determination of the endpoint where the hammer should make a sound in the video, thereby interfering with their temporal integration and recalibration of the non-speech audiovisual stimulus. It cannot be ruled out that the later maturation of non-speech stimuli than speech stimuli is potentially attributed to deficiencies in task materials as mentioned above, necessitating future research to validate this finding through task refinement.

The relationship between rapid audiovisual temporal recalibration and reading

The present study found a significant correlation between rapid audiovisual temporal recalibration and reading fluency, while rapid audiovisual temporal recalibration could explain the unique variance of reading fluency. It is worth noting that irrespective of the type of stimuli (speech vs. non-speech) processed and the age group of participants (children vs. adults), rapid audiovisual temporal recalibration could significantly predict reading ability. One possible explanation for this effect is that rapid temporal recalibration can generate prior knowledge through continuously receiving new temporal relationships (e.g., different SOA between audiovisual stimuli), thereby affecting the perception of temporal sequence (perception of simultaneity, succession, and sequence of objective events; Sato & Aihara, 2011). Moreover, the increase in the TBW in the initial stage of temporal recalibration also reflects that the brain can mitigate discomfort caused by successive sensory stimuli by reducing the resolution of the perception of temporal sequence, thus facilitating the integration of perception and memory (Yuan & Huang, 2011). The temporal sequential processing of arranging the order of characters and their sounds is a crucial aspect of reading (Vidyasagar & Pammer, 2010), as it enables individuals to perceive character positions and overall glyphs, thereby promoting their acquisition of grapheme-phoneme correspondences (Hood & Conlon, 2004). Therefore, the stronger an individual’s ability to rapid temporal recalibration, the stronger their temporal sequence perception will be, thereby promoting the development of a series of reading-related skills such as graphemes-phonemes bundling and ultimately enabling fluent reading.

Another possible explanation for this effect is that, within the Bayesian framework, the brain generates an estimation of the reliability of the information from each sensory channel by perceiving the interrelationship between information derived from different sensory channels. Subsequently, it utilizes this prior knowledge to compare with stimuli inputted through the current sensory channels and adjusts the weighting of each channel’s information during multisensory integration, to optimize decision-making (Ernst & Bühlhoff, 2004). Temporal recalibration can be regarded as a dynamic process that adapts to varying temporal delays. Within a dynamic context, prior knowledge regarding the temporal relationship is re-established, subsequently affecting the multisensory integration under different temporal relationships (Noel *et al.*, 2016; Sato & Aihara, 2011).

Specifically, the existing prior knowledge can produce top-down predictions (i.e., the internal sensory representation). The recalibration process aims to minimize the discrepancy between these top-down predictions and actual sensory inputs, thereby optimizing the multisensory temporal integration (Noel et al., 2017). An appropriate audiovisual TBW is important for the proper binding of speech sounds to print when learning to read (Gori et al., 2020). Given the significant correlation between rapid audiovisual temporal recalibration and audiovisual temporal integration (Harvey et al., 2014; Noel et al., 2016; van der Burg et al., 2013), a strong ability of rapid audiovisual temporal recalibration can improve reading ability by flexibly modulating temporal window. Previous studies have found that individuals with dyslexia exhibited phonetic recalibration deficits at the verbal level (Keetels et al., 2018). Moreover, their rapid neural adaptation ability was comparatively weaker than that of typical readers (Perrachione et al., 2016), and they were unable to effectively integrate audiovisual temporal information according to the optimal Bayesian principle (Gori et al., 2020). These aforementioned findings collectively provide evidence that rapid audiovisual temporal recalibration can effectively affect reading ability.

Furthermore, our results revealed that the rapid audiovisual temporal recalibration (i.e., Δ PSS) could better explain the variance of reading ability than the average indicator of audiovisual temporal integration (i.e., TBW). This indicates that rapid audiovisual temporal recalibration is a more effective indicator for investigating reading ability. Some previous studies have reported no significant differences in TBW or related indicators between dyslexic readers with impaired reading ability and typically developing readers (e.g., Francisco et al., 2014; Francisco, Groen, et al., 2017; Francisco, Jesse, et al., 2017; Laasonen et al., 2002). However, Noel et al. (2017) found that individuals with autism spectrum disorders (ASD), who exhibited significantly lower scores on vocabulary tests compared to typically developing individuals, had no significant difference in TBW when integrating non-speech audiovisual stimuli from typically developing individuals. Nevertheless, their Δ PSS (representing the ability of rapid audiovisual temporal recalibration) was significantly smaller than that of typically developing individuals. Consequently, our results also provide a plausible explanation for some previous studies on dyslexia that failed to find that the average level of audiovisual temporal integration is related to reading ability (e.g., Francisco et al., 2014; Francisco, Groen, et al., 2017; Francisco, Jesse, et al., 2017; Laasonen et al., 2002).

Limitations and prospects

The present study has several limitations: (1) The present study exclusively only focused on children approximately 10 years old. Future research should expand this investigation to the entire span of childhood, to delineate the developmental trajectory of rapid audiovisual temporal recalibration ability throughout childhood and establish comparisons with adults. (2) Only two types of stimuli were included in our experiments. It is worthwhile to investigate more simple non-speech stimuli (e.g., flash-beep stimuli) or more complex speech-related stimuli (e.g., words, sentences), which may exhibit distinct developmental trajectories. (3) Considering that the audiovisual SJ task requires monotonous and continuous button pressing

for a duration of 10 ~ 12 minutes, the present study incorporated stimulus type (speech vs. non-speech) as a between-subject factor to prevent children from experiencing boredom and responding insincerely. Future studies may benefit from including stimulus type as a within-subject factor to facilitate a more comprehensive comparison between the speech and non-speech stimuli (e.g., the interaction of age and stimulation type). (4) Notice that our main goal was to primarily reveal the relationship between rapid audiovisual temporal recalibration and reading ability. It was beyond our scope to clarify the causal inferences, which is a critical question for future research. The investigation into whether individuals with dyslexia exhibit deficits in rapid audiovisual temporal recalibration and whether training this ability can improve the reading ability of dyslexic readers will contribute to addressing this research question.

Conclusions

The present study found that children, approximately 10 years old, exhibit a strong rapid audiovisual temporal recalibration ability for speech and non-speech stimuli. In the case of speech stimuli, no significant differences were observed between children and adults regarding Δ TBW and Δ PSS. Nevertheless, the trial-by-trial recalibration process of rapid audiovisual temporal recalibration in children exhibits potential for further development. For non-speech stimuli, significant differences between children and adults were observed in both the Δ PSS indicator and the trial-by-trial recalibration process. Moreover, the rapid audiovisual temporal recalibration ability for both speech and non-speech stimuli explained reading fluency uniquely in both children and adults.

Supplementary material. The supplementary material for this article can be found at <https://doi.org/10.1017/S0142716425000098>

Replication package. Study materials, data, and analysis codes are available at <https://doi.org/10.6084/m9.figshare.25534639.v2>.

Competing interests. The authors declare that no competing interests exist.

Ethics approval statement. This study was approved by the Ethics Committee of the Institute of Psychology, Chinese Academy of Science.

References

- Alsius, A., Navarra, J., Campbell, R., & Soto-Faraco, S. (2005). Audiovisual integration of speech falters under high attention demands. *Current Biology*, *15*(9), 839–843.
- Beneventi, H., Tønnessen, F. E., Erslund, L., & Hugdahl, K. (2010). Executive working memory processes in dyslexia: behavioral and fMRI evidence. *Scandinavian Journal of Psychology*, *51*(3), 192–202.
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289–300.
- Birch, H. G., & Belmont, L. (1964). Auditory-visual integration in normal and retarded readers. *American Journal of Orthopsychiatry*, *34*(5), 852–861.
- Burr, D., & Gori, M. (2012). Multisensory integration develops late in humans. In M. M. Murray & M. T. Wallace (Eds.), *The Neural Bases of Multisensory Processes* (pp. 345–362). Boca Raton, FL: CRC Press.

- Chein, J. M., & Schneider, W.** (2005). Neuroimaging studies of practice-related change: fMRI and meta-analytic evidence of a domain-general control network for learning. *Cognitive Brain Research*, *25*(3), 607–623.
- Chen, Y. C., Shore, D. I., Lewis, T. L., & Maurer, D.** (2016). The development of the perception of audiovisual simultaneity. *Journal of Experimental Child Psychology*, *146*, 17–33.
- De Niar, M. A., Noel, J. P., & Wallace, M. T.** (2017). The impact of feedback on the different time courses of multisensory temporal recalibration. *Neural Plasticity*, *2017*, 3478742.
- Dijkstra, T., Schreuder, R., & Frauenfelder, U. H.** (1989). Grapheme context effects on phonemic processing. *Language and Speech*, *32*(2), 89–108.
- Dixon, N. F., & Spitz, L.** (1980). The detection of auditory visual desynchrony. *Perception*, *9*(6), 719–721.
- Ernst, M. O., & Bühlhoff, H. H.** (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences*, *8*(4), 162–169.
- Francisco, A. A., Groen, M. A., Jesse, A., & McQueen, J. M.** (2017). Beyond the usual cognitive suspects: the importance of speechreading and audiovisual temporal sensitivity in reading ability. *Learning and Individual Differences*, *54*, 60–72.
- Francisco, A. A., Jesse, A., & McQueen, J. M.** (2014). Audiovisual temporal sensitivity in typical and dyslexic adult readers. In *Interspeech 2014: 15th Annual Conference of the International Speech Communication Association* (pp. 2575–2579). Munich: Max Planck Institute.
- Francisco, A. A., Jesse, A., Groen, M. A., & McQueen, J. M.** (2017). A general audiovisual temporal processing deficit in adult readers with dyslexia. *Journal of Speech, Language, and Hearing Research*, *60*(1), 144–158.
- Fujisaki, W., Shimojo, S., Kashino, M., & Nishida, S. Y.** (2004). Recalibration of audiovisual simultaneity. *Nature Neuroscience*, *7*(7), 773–778.
- Gori, M., Ober, K. M., Tinelli, F., & Coubard, O. A.** (2020). Temporal representation impairment in developmental dyslexia for unisensory and multisensory stimuli. *Developmental Science*, *23*, e12977.
- Hairston, W. D., Burdette, J. H., Flowers, D. L., Wood, F. B., & Wallace, M. T.** (2005). Altered temporal profile of visual-auditory multisensory interactions in dyslexia. *Experimental Brain Research*, *166*(3–4), 474–480.
- Harvey, C., van der Burg, E., & Alais, D.** (2014). Rapid temporal recalibration occurs crossmodally without stimulus specificity but is absent unimodally. *Brain Research*, *1585*, 120–130.
- Hillock-Dunn, A., Grantham, D. W., & Wallace, M. T.** (2016). The temporal binding window for audiovisual speech: children are like little adults. *Neuropsychologia*, *88*, 74–82.
- Hood, M., & Conlon, E.** (2004). Visual and auditory temporal processing and early reading development. *Dyslexia*, *10*(3), 234–252.
- Keetels, M., Bonte, M., & Vroomen, J.** (2018). A selective deficit in phonetic recalibration by text in developmental dyslexia. *Frontiers in Psychology*, *9*, 710.
- Koelewijn, T., Bronkhorst, A., & Theeuwes, J.** (2010). Attention and the multiple stages of multisensory integration: a review of audiovisual studies. *Acta Psychologica*, *134*, 372–384.
- Laasonen, M., Service, E., & Virsu, V.** (2002). Crossmodal temporal order and processing acuity in developmentally dyslexic young adults. *Brain and Language*, *80*(3), 340–354.
- Lewkowicz, D. J.** (1996). Perception of auditory-visual temporal synchrony in human infants. *Journal of Experimental Psychology: Human Perception and Performance*, *22*(5), 1094–1106.
- Lewkowicz, D. J., & Flom, R.** (2014). The audiovisual temporal binding window narrows in early childhood. *Child Development*, *85*, 685–694.
- Liu, S., Wang, L. C., & Liu, D.** (2019). Auditory, visual, and cross-modal temporal processing skills among Chinese children with developmental dyslexia. *Journal of Learning Disabilities*, *52*(6), 431–441.
- Meng, Z. L., Liu, M. L., & Bi, H.-Y.** (2022). Spatial and temporal processing difficulties in Chinese children with developmental dyslexia: an ERP study. *Dyslexia*, *28*(4), 416–430.
- Meng, Z. L., Wydell, T. N., & Bi, H.-Y.** (2019). Visual-motor integration and reading Chinese in children with/without dyslexia. *Reading and Writing*, *32*(2), 493–510.
- Mossbridge, J., Zweig, J., Grabowecy, M., & Suzuki, S.** (2017). An association between auditory-visual synchrony processing and reading comprehension: behavioral and electrophysiological evidence. *Journal of Cognitive Neuroscience*, *29*(3), 435–447.
- Murray, M. M., Lewkowicz, D. J., Amedi, A., & Wallace, M. T.** (2016). Multisensory processes: a balancing act across the lifespan. *Trends in Neurosciences*, *39*(8), 567–579.

- Navarra, J., Hartcher-O'Brien, J., Piazza, E., & Spence, C. (2009). Adaptation to audiovisual asynchrony modulates the speeded detection of sound. *Proceedings of the National Academy of Sciences*, **106**(23), 9169–9173.
- Navarra, J., Soto-Faraco, S., & Spence, C. (2007). Adaptation to audiotactile asynchrony. *Neuroscience Letters*, **413**(1), 72–76.
- Navarra, J., Vatakis, A., Zampini, M., Soto-Faraco, S., Humphreys, W., & Spence, C. (2005). Exposure to asynchronous audiovisual speech extends the temporal window for audiovisual integration. *Cognitive Brain Research*, **25**(2), 499–507.
- Noel, J. P., De Nier, M. A., Stevenson, R., Alais, D., & Wallace, M. T. (2017). Atypical rapid audio-visual temporal recalibration in autism spectrum disorders. *Autism Research*, **10**(1), 121–129.
- Noel, J. P., De Nier, M. A., van der Burg, E., & Wallace, M. T. (2016). Audiovisual simultaneity judgment and rapid recalibration throughout the lifespan. *PLoS One*, **11**(8), e0161698.
- Pammer, K., & Vidyasagar, T. R. (2005). Integration of the visual and auditory networks in dyslexia: a theoretical perspective. *Journal of Research in Reading*, **28**(3), 320–331.
- Perrachione, T. K., Del Tufo, S. N., Winter, R., Murtagh, J., Cyr, A., Chang, P., Halverson, K., Ghosh, S., Christodoulou, J., & Gabrieli, J. D. (2016). Dysfunction of rapid neural adaptation in dyslexia. *Neuron*, **92**(6), 1383–1397.
- Powers, A. R., Hevey, M. A., & Wallace, M. T. (2012). Neural correlates of multisensory perceptual learning. *Journal of Neuroscience*, **32**(18), 6263–6274.
- Powers, A. R., Hillock, A. R., & Wallace, M. T. (2009). Perceptual training narrows the temporal window of multisensory binding. *Journal of Neuroscience*, **29**(39), 12265–12274.
- Qian, Y., Deng, Y., Zhao, J., & Bi, H. Y. (2015). Magnocellular-dorsal pathway function is associated with orthographic but not phonological skill: fMRI evidence from skilled Chinese readers. *Neuropsychologia*, **71**, 84–90.
- Sato, Y., & Aihara, K. (2011). A Bayesian model of sensory adaptation. *PLoS One*, **6**(4), e19377.
- Stevenson, R. A., & Wallace, M. T. (2013). Multisensory temporal integration: task and stimulus dependencies. *Experimental Brain Research*, **227**(2), 249–261.
- Stevenson, R. A., Fister, J. K., Barnett, Z. P., Nidiffer, A. R., & Wallace, M. T. (2012). Interactions between the spatial and temporal stimulus factors that influence multisensory integration in human performance. *Experimental Brain Research*, **219**(1), 121–137.
- Tanaka, A., Koizumi, A., Imai, H., Hiramatsu, S., Hiramoto, E., & de Gelder, B. (2010). I feel your voice: cultural differences in the multisensory perception of emotion. *Psychological Science*, **21**(9), 1259–1262.
- Theves, S., Chan, J. S., Naumer, M. J., & Kaiser, J. (2020). Improving audio-visual temporal perception through training enhances beta-band activity. *NeuroImage*, **206**, 116312.
- van der Burg, E., & Goodbourn, P. T. (2015). Rapid, generalized adaptation to asynchronous audiovisual speech. *Proceedings of the Royal Society B: Biological Sciences*, **282**(1804), 20143083.
- van der Burg, E., Alais, D., & Cass, J. (2013). Rapid recalibration to audiovisual asynchrony. *Journal of Neuroscience*, **33**, 14633–14637.
- van der Burg, E., Alais, D., & Cass, J. (2015). Audiovisual temporal recalibration occurs independently at two different time scales. *Scientific Reports*, **5**, 14526.
- van der Burg, E., Olivers, C. N. L., Bronkhorst, A. W., & Theeuwes, J. (2008). Pip and pop: nonspatial auditory signals improve spatial visual search. *Journal of Experimental Psychology: Human Perception and Performance*, **34**(5), 1053–1065.
- van Wassenhove, V., Grant, K. W., & Poeppel, D. (2007). Temporal window of integration in auditory-visual speech perception. *Neuropsychologia*, **45**(3), 598–607.
- Vercoulen, J. H., Swanink, C. M., Fennis, J. F., Galama, J. M., van der Meer, J. W., & Bleijenberg, G. (1994). Dimensional assessment of chronic fatigue syndrome. *Journal of Psychosomatic Research*, **38**(5), 383–392.
- Vidyasagar, T. R., & Pammer, K. (2010). Dyslexia: a deficit in visuo-spatial attention, not in phonological processing. *Trends in Cognitive Sciences*, **14**(2), 57–63.
- Virsu, V., Lahti-Nuutila, P., & Laasonen, M. (2003). Crossmodal temporal processing acuity impairment aggravates with age in developmental dyslexia. *Neuroscience Letters*, **336**(3), 151–154.
- Vroomen, J., & Keetels, M. (2010). Perception of intersensory synchrony: a tutorial review. *Attention, Perception, and Psychophysics*, **72**(4), 871–884.

- Wu, H. D.** (2020). *Visual, Auditory and Cross-Modal Temporal Sensitivity of Chinese Children with Developmental Dyslexia and Their Effects on Reading Abilities*. Shanghai: East China Normal University.
- Yuan, X. Y., & Huang, X. T.** (2011). Temporal recalibration in multisensory integration. *Advances in Psychological Science*, *19*(5), 692–700.
- Zaidel, A., Goin-Kochel, R. P., & Angelaki, D. E.** (2015). Self-motion perception in autism is compromised by visual noise but integrated optimally across multiple senses. *Proceedings of the National Academy of Sciences*, *112*(20), 6461–6466.
- Zhou, H. Y., Shi, L. J., Yang, H. X., Cheung, E. F., & Chan, R. C.** (2020). Audiovisual temporal integration and rapid temporal recalibration in adolescents and adults: age-related changes and its correlation with autistic traits. *Autism Research*, *13*(4), 615–626.