The Role of Letter–Speech Sound Integration in Native and Second Language Reading: A Study in Native Japanese Readers Learning English

Dongyang Yan and Ayumi Seki

Abstract

■ The automatic activation of letter–speech sound (L-SS) associations is a vital step in typical reading acquisition. However, the contribution of L-SS integration during nonalphabetic native and alphabetic second language (L2) reading remains unclear. This study explored whether L-SS integration plays a similar role in a nonalphabetic language as in alphabetic languages and its contribution to L2 reading among native Japanese-speaking adults with varying English proficiency. A priming paradigm in Japanese and English was performed by presenting visual letters or symbols, followed by auditory sounds. We compared behavioral and event-related responses elicited by congruent letter–sound pairs, incongruent pairs, and baseline condition (symbol–sound pairs). The behavioral experiment revealed shorter RTs in the congruent condition for Japanese and English tasks, suggesting a facilitation effect of congruency. The ERP experiment results showed an increased early N1 response to Japanese congruent pairs compared to corresponding incongruent stimuli at the left frontotemporal electrodes. Interestingly, advanced English learners exhibited greater activities in bilateral but predominantly right-lateralized frontotemporal regions for the congruent condition within the N1 time window. Moreover, the enhancement of P2 response to congruent pairs was observed in intermediate English learners. These findings indicate that, despite deviations from native language processing, advanced speakers may successfully integrate letters and sounds during English reading, whereas intermediate learners may encounter difficulty in achieving L-SS integration when reading L2. Furthermore, our results suggest that L2 proficiency may affect the level of automaticity in L-SS integration, with the right P2 congruency effect playing a compensatory role for intermediate learners. ■

INTRODUCTION

Reading, as one way of acquiring knowledge and information, is an important method to successfully make academic and social achievements. In an alphabetic writing system, words are composed of sets of graphemes (letters) that correspond to the phonemes (speech sounds). Becoming a reading expert heavily hinges on the automatic activation of the established letter–speech sound (L-SS) association (Ehri, 2005). In most natural settings, typically developing children need to converge a letter with a speech sound into a new audiovisual (AV) unit, and it takes a considerably long time for them to automatically extract specific information from this L-SS association (e.g., Karipidis et al., 2017, 2018). As a vital part of learning to read, the cross-modal integration of letters and speech sounds serves as the fundamental starting point for building an effective brain network that facilitates fluent reading. Hence, the investigation of L-SS integration is essential for typical reading acquisition not only in the native language (L1) but also in a second language (L2). Yet, learning an L2 is exceedingly complicated and challenging, as there is a significant influence of the learner's L1 background on L2 acquisition, which may make L-SS integration contribute differently to alphabetic L2 reading.

The general features of AV integration processing (e.g., lipreading, which involves integration of lip movements and speech sounds) have been widely explored by manipulating the congruency of different AV inputs. In this paradigm, two conditions have been presented: (1) congruent auditory and visual inputs and (2) mismatching inputs (i.e., incongruent AV pairs). Murray and Wallace (2011) proposed that when auditory and visual stimuli are congruent, their interaction can accelerate and improve the ability in perception. In line with AV speech integration, this simple framework has also been used in L-SS integration studies. Using this congruency effect, an early behavioral priming study conducted on Dutch adults showed faster RTs in identifying a vowel in an auditory syllable (e.g., /a/ in /ka/) presented after a congruent visual letter prime (Dijkstra, Schreuder, & Frauenfelder, 1989). More recently, a magnetoencephalography study by Raij, Uutela, and Hari (2000) compared the activation elicited by auditory (speech sounds), visual (letters), and AV (congruent or conflicting) stimuli. Successful L-SS integration was found to correspond to areas, including the superior temporal cortex (STC), that served as a heteromodal AV

Faculty of Education, Hokkaido University, Japan

Journal of Cognitive Neuroscience 36:6, pp. 1123–1140
https://doi.org/10.1162/jocn_a_02137

integration site. Later functional neuroimaging studies of L-SS integration (Blau, Van Atteveldt, Formisano, Goebel, & Blomert, 2008; van Atteveldt, Formisano, Blomert, & Goebel, 2007; van Atteveldt, Formisano, Goebel, & Blomert, 2007; Herdman et al., 2006; van Atteveldt, Formisano, Goebel, & Blomert, 2004) have reported significantly greater activation for congruent conditions than for incongruent conditions (i.e., congruency effect) in multimodal areas, the STC, and sensory-specific areas including the auditory (i.e., Heschl's sulcus, planum temporale) and visual cortices. In light of these findings, van Atteveldt, Roebroeck, and Goebel (2009) proposed the feedback hypothesis, suggesting that the unisensory signals of multisensory stimuli are initially integrated in the STC, and then the feedback input from the STC would modulate activities in the auditory cortex. Therefore, one expects the congruency effects to be mainly observed in the heteromodal integration area and the auditory cortex.

With the high time resolution feature, neurophysiological studies including EEG have been performed to explore the time window of multisensory integration processing. In a series of ERP studies, the mismatch negativity (MMN) component was used to investigate L-SS integration (Žarić et al., 2014; Andres, Oram Cardy, & Joanisse, 2011; Froyen, Willems, & Blomert, 2011; Froyen, Bonte, van Atteveldt, & Blomert, 2009; Froyen, Van Atteveldt, Bonte, & Blomert, 2008). The MMN is elicited by a rare AV pair (visual "a," auditory /o/; incongruent) that is presented in a sequence of frequent stimuli (visual "a," auditory /a/; congruent). These studies have demonstrated the bimodal enhancement of the MMN in typical readers, occurring approximately 100–250 msec after the onset of the auditory stimulus that appears synchronously with the corresponding letter (AV0) and 200 msec later (AV200) than the presence of the visual letter, when compared to that of the MMN evoked by the auditory-only condition. This component has been considered as the completion of early and automatic integration (see a review from Fitzgerald & Todd, 2020). However, there may be some limitations when using MMN as a reliable indicator of L-SS integration. First, different from the theory of feedback hypothesis, the enhancement of MMN amplitude might be attributed to incongruent trials, and the observed incongruency effects are more likely to stem from neural adaptation (Grill-Spector & Malach, 2001) or cross-modal repetition (Henson, 2003). In addition, considering that many different processes presumably overlap within the 100- to 250-msec time window, it might be difficult to draw firm conclusions on the precise time course of L-SS integration within this time range. Therefore, relying solely on MMN as an indicator may not be appropriate to precisely investigate the temporal dynamics of L-SS integration.

Empirical studies have proposed that auditory ERP components are modulated by the presence of the congruent visual context. Earlier studies investigating AV integration (e.g., lipreading) have reported amplitude differences in brain responses occurring 100–250 msec after the onset of congruent auditory stimuli, compared to that for incongruent inputs (e.g., Roa Romero, Senkowski, & Keil, 2015; Kaganovich & Schumaker, 2014; Knowland, Mercure, Karmiloff-Smith, Dick, & Thomas, 2014; Stekelenburg & Vroomen, 2007). These brain responses are typically represented by the auditory N1 and P2 ERP components. The N1/P2 components, with maximal current distribution at the central scalp sites (i.e., at electrode Cz), often referred to together as the vertex potential, are highly responsive to auditory speech (Hoonhorst et al., 2009; Pang & Taylor, 2000). Other related studies have shown that the amplitude difference between matching and mismatching stimuli in the auditory N1 component appears as early as 100 msec (Kaganovich & Schumaker, 2014; Van Wassenhove, Grant, & Poeppel, 2005; Klucharev, Möttönen, & Sams, 2003) and a shorter latency of N1 when auditory speech is accompanied by the consistent lipreading information (Knowland et al., 2014; Bernstein, Auer, Wagner, & Ponton, 2008; van Wassenhove et al., 2005). In contrast, only a few ERP studies have explored the time course of L-SS integration through the congruency effect, as an indicator of cross-modal facilitation or AV integration. Recently, Nash et al. (2017) used auditory N1/P2 ERP components to explore the L-SS integration in English children with and without dyslexia. They utilized a priming paradigm by presenting visual letters 1000 msec before the onset of letter sounds. Their results supported the feedback hypothesis, showing stronger congruency effects at the left frontotemporal electrodes within 50–125 msec for typical advanced readers, whereas a congruency effect in the younger group with lower reading fluency was observed in the centroparietal regions within approximately 200–300 msec at the P2 time span. These findings suggest that the impact of reading fluency on the time course of integrating letters and speech sounds shifts from a late to an early congruency effect on ERP amplitude as reading fluency improves. Meanwhile, the results from Nash et al. (2017) tended to concur with studies focused on reading development, highlighting the AV congruency effect observed at the basic phoneme level, which involves L-SS pairs (Caffarra, Lizarazu, Molinaro, & Carreiras, 2021; Karipidis et al., 2017, 2018), as well as at the word level (Eberhard-Moscicka, Jost, Daum, & Maurer, 2021; Jost, Eberhard-Moscicka, Frisch, Dellwo, & Maurer, 2014). These studies collectively indicate that the AV congruency effect on ERP amplitude may serve as a strong predictor of reading fluency.

However, it remains largely unclear about the contribution of L-SS integration to L2 reading, particularly when the native language and the learned L2 have different orthography–phonology mappings. In alphabetic languages, literacy acquisition typically begins with learning L-SS associations. However, the writing system in Japanese is substantially different from that of alphabetic languages, as it combines a mixture of both semantic symbols known as "kanji" and orthographically consistent phonetic symbols known as *kana.* In addition, the phonological units in Japanese orthography, known as *mora,* are larger than the phonemes in English. Furthermore, it is important to note that the level of L2 proficiency might be one critical factor impacting L2 AV integration. Neuroimaging studies have reported that the native language background may influence brain responses to L2 reading, as well as the L2 proficiency and age of acquisition (Bowden, Steinhauer, Sanz, & Ullman, 2013; Perani & Abutalebi, 2005; Nakada, Fujii, & Kwee, 2001). These studies indicate that the brain response to an L2 may show greater similarity to the response to their native language among more proficient learners compared to those with lower proficiency levels and that highly proficient learners are more likely to develop L2 processing that resemble those found in the native language processing. More related MMN studies by Wang et al. (2019, 2022) have examined L-SS integration in L2 reading among native Chinese speakers with varying L2 proficiency. They found that local native Chinese speakers failed to integrate English letters and sounds, in contrast to their native English peers, whereas native Chinese speakers with English immersion experience successfully integrated English letters and sounds like native English speakers. In light of these studies, we investigated whether the English proficiency of native Japanese speakers affects the role of L-SS integration in L2 reading.

The aim of the current study was to testify the role of L-SS integration in nonalphabetic native and alphabetic L2 reading using behavioral and ERP measurements in native Japanese-speaking adults with various English proficiency. To achieve this, we employed a priming paradigm adapted from Nash et al. (2017). Previous studies using this paradigm (Clayton & Hulme, 2018; Nash et al., 2017) have found that both typical English readers and dyslexic children exhibited facilitation of the congruency effect in their behavioral responses, but the engaged neural circuitry differed. On this account, it is crucial to examine both behavioral and neural responses in L2 readers. To adapt the paradigm to adults, we amended the SOA by presenting the visual stimuli leading the auditory stimuli by 200 msec. This adjustment was informed by prior research suggesting a smaller window size for AV integration in typically developing adults (Zhou, Cheung, & Chan, 2020; Wallace & Stevenson, 2014). The choice of the 200 msec as the SOA was based on previous studies (e.g., Wang et al., 2019; Froyen et al., 2008). In line with the congruency effects from previous studies, we hypothesized that the congruency effects of early ERP response should be found when a native Japanese reads *kana*, transparent phonetic symbols. Regarding English reading, given English with a more opaque orthography, we expected that the formation of letter–sound associations would require a longer learning time. Hence, we predicted that advanced English (AE) learners would demonstrate a greater congruency effect in the early auditory response, whereas intermediate English (IE) learners would not.

METHODS

Participants

Thirty-six native Japanese-speaking university students (age range = 18–27 years; 20 women) were recruited to participate in this study. All the participants passed a rigorous examination in Japanese, a prerequisite for the university admission process, suggesting a high level of Japanese proficiency among our participants. The sample size for this study was determined based on the previous study by Nash et al. (2017). Considering the impact of age on L2 acquisition (Cao, Tao, Liu, Perfetti, & Booth, 2013), we recruited participants who started to learn English after the entry to primary school. Data of two female participants who failed to complete the EEG experiment were excluded. The remaining 34 participants took part in both behavioral and EEG experiments. Procedure of the study was approved by the ethics committee of the Department of Education at Hokkaido University and conformed to the Declaration of Helsinki. Written informed consent was gained from all the participants. All the participants were right-handed and had normal or corrected-to-normal vision. In addition, they had normal hearing and no record of reading difficulties diagnoses, or other neurological or psychiatric disorders.

Procedure

All participants were asked to complete an online questionnaire before attending the laboratory for the study. Behavioral and EEG data were collected on two separate days. Participants were informed about the experimental procedure and were introduced to the facilities before the conduction of the experiments. On Day 1, they were instructed to finish a pretest (70 min), which included a batch of standardized measures of English proficiency and general cognitive ability, and a behavioral letter–sound priming experiment (20 min). The priming experiment while EEG data were recording was performed on a different day, typically within a 10-day interval after the initial measurement that included the pretest and the behavioral priming task.

Questionnaire

The online questionnaire was used to assess the participants' language backgrounds and attitudes toward L2 learning. We made the questionnaire based on the language history questionnaire for speakers with multiple languages by Li, Zhang, Yu, and Zhao (2020), but with fewer questions. This is because our questionnaire focused more on English learning experience, their English usage as an L2, and their age of the English acquisition. The first part of the questionnaire included general information questions about the participants, such as age, gender, L2 language proficiency, and English proficiency test scores (e.g., TOEFL). Subsequently, the following section checked participants' overseas experiences and explored their L2 usage when living overseas. Afterward, the responses to questions about their attitudes toward English learning were collected through the questionnaire.

L2 Proficiency and General Cognitive Measures

We used three standardized tests of English literacy to measure participants' basic reading abilities in English.

Test of Word Reading Efficiency II: Sight Word Efficiency and Phonetic Decoding Efficiency

For the Test of Word Reading Efficiency II (Torgeson, Wagner, & Rashotte, 1999), participants were instructed to read aloud a list of 108 words and 66 pronounceable printed nonwords accurately and quickly, with a time limit of 45 sec, respectively.

Woodcock–Johnson III: Letter–Word Identification, Word Attack, Reading Fluency, and Spelling of Sounds

For the Letter–Word Identification (76 words) and Word Attack (32 pseudowords) tests in Woodcock–Johnson III (Schrank, 2010), participants were asked to provide the correct pronunciation of these words and pseudowords. There is no time limit, but they needed to read fluently and smoothly. Each test was stopped after six consecutive errors or no responses.

The Reading Fluency test, with a 3-min time limit, was used to assess English (automatic) semantic reading ability and generic knowledge. Participants judged each short sentence as truth or paradox by accurately and quickly marking "Yes" or "No," with a maximum of 98 questions.

Spelling ability was measured using the Spelling of Sounds test. The audio recording of pronounceable nonwords was played twice, and participants were asked to write down these heard nonwords correctly. A maximum of 28 nonwords was provided. The test was stopped after four consecutive errors or no responses.

Comprehensive Test of Phonological Processing–Second Edition: Elision, Blending Words, Blending Nonwords, and Segmenting Nonwords

The Comprehensive Test of Phonological Processing–Second Edition (Dickens, Meisinger, & Tarar, 2015) was used to assess reading-related phonological processing skills. Participants removed or segmented phonemes to form other words or pronounceable nonwords in the Elision test and Segmenting Nonwords test. In the Blending Words and Blending Nonwords tests, participants were asked to synthesize sounds to form words or nonwords. All four subtests were terminated after three consecutive errors or no responses.

Matrix Reasoning Test

General cognitive ability was assessed using a nonverbal intelligence test from the Wechsler Abbreviated Scale of Intelligence (Wechsler, 1999). A matrix reasoning test was used to measure the participants' ability to solve new problems, perceive relationships, and complete visual analogies without evaluating their vocabulary or language proficiency. The test consists of 26 questions and was stopped after four consecutive errors or no responses.

Given the strong correlation observed among the English proficiency tests, it is difficult to isolate one specific test to serve as the exclusive criterion for grouping participants. Hence, to establish the grouping criteria, principal component analysis was performed on all the English proficiency test scores to extract the principal components related to the grouping criterion. This analysis allowed us to identify the principal component that accounted for the highest proportion of the variance. Accordingly, the first principal component accounts for 56.67% of the overall variability, indicating comprehensive English proficiency. The second component with a contribution of 11.08% likely explains aspects of phonological and phonemic awareness (see Appendix for a summary of the principal component analysis results). Consequently, we utilized the first principal component as the primary criterion for grouping participants. Thirty-four participants were classified into two groups based on the results of their behavioral English proficiency tests: an AE group ($n = 13$) and an IE group ($n = 21$). The demographic information and behavioral tests for each group, including age, gender, and performance of all behavioral tasks, are summarized in Table 1. There was no difference between two groups in terms of their mean scaled scores on the Matrix Reasoning Test. Notably, it is worth mentioning that a significant portion of participants in the AE group had passed English proficiency tests for application to U.S. universities (e.g., TOEFL) and had resided in English-speaking countries for more than 6 months (9 of 13 participants).

Behavioral Priming Measure

A priming experimental paradigm adapted from Nash et al. (2017) was performed in both Japanese and English. To be consistent with previous studies, we used both incongruent and baseline conditions. There were five conditions for each language task, as shown in Figure 1A.

In congruent and incongruent AV conditions, the prime and the target were matching and mismatching kana/letters and sounds (e.g., kana "と" or letter "k" and sound /to/ or /k/). In the baseline condition, an International Phonetic Alphabet letter was used as the visual prime (e.g., symbol "ʉ"). The International Phonetic Alphabet letter is an adapted alphabetic letter based on the Latin script that is unfamiliar to the Japanese participants. The original speech sounds were recorded by male native speakers at 44.1-kHz, 16-bit quantization. We then

Table 1. Demographic Information and Mean (*SD*) of Scores on Behavioral English Proficiency and General Cognitive Ability in the Two Groups

Measure (Points)	*AE Learners* *Mean (SD)*	*IE Learners* *Mean (SD)*	*Significance Testing*	*Group Differences*
N (men/women)	13 (4/9)	21 (12/9)	–	–
Age (years)	20.54 (2.56)	21.52 (2.20)	p = .256	AE = IE
Age (years) of L2 acquisition	10.54 (3.52)	11.33 (2.59)	p = .469	AE = IE
Number of individuals learning English out of interest (%)	10 (76.92%)	10 (47.62%)	p = .092	AE = IE
Length of time since the start of English acquisition (years)	10.00 (2.38)	10.19 (3.16)	p = .853	AE = IE
TOWRE II Sight Word Efficiency raw score	82.69 (6.82)	71.52 (4.89)	p = .000	AE > IE
TOWRE II Phonemic Decoding Efficiency raw score	47.54 (4.20)	35.29 (6.08)	p = .000	AE > IE
WJ III LWI raw score	55.15 (1.91)	50.90 (3.06)	p = .000	AE > IE
WJ III WA raw score	28.46 (2.03)	24.76 (2.55)	p = .000	AE > IE
WJ III RF raw score	68.77 (12.04)	44.90 (7.12)	p = .000	AE > IE
WJ III SS raw score	22.54 (2.60)	19.14 (2.89)	p = .002	AE > IE
CTOPP-2 Elision raw score	31.00 (1.63)	27.62 (4.85)	p = .022	AE > IE
CTOPP-2 Blending Words raw score	23.46 (1.66)	17.71 (5.88)	p = .002	AE > IE
CTOPP-2 Blending Nonwords raw score	15.77 (2.68)	10.86 (4.35)	p = .001	AE > IE
CTOPP-2 Segmenting Nonwords raw score	21.46 (2.79)	18.29 (3.58)	p = .010	AE > IE
Matrix Reasoning Test raw score	21.46 (1.76)	21.71 (2.03)	p = .713	AE = IE

TOWRE = Test of Word Reading Efficiency; WJ = Woodcock–Johnson; LWI = Letter–Word Identification; WA = Word Attack; RF = Reading Fluency; SS = Spelling of Sounds; CTOPP-2 = Comprehensive Test of Phonological Processing–Second Edition.

off-line down-sampled these files to 22.05 kHz and matched loudness at 70 dB using the PRAAT software (www.fon.hum.uva.nl/praat/; Boersma & Weenink, 2020).

We included two control conditions to balance the button press and to avoid the condition that only the "yes" responses to speech sounds were involved. In the two control conditions, the prime was either a kana/letter or a symbol, but the auditory stimulus was a scrambled phoneme edited from the corresponding speech phonemes using MATLAB (The MathWorks) code from Ellis (2010). We chopped each speech sound into 5-msec overlapping Hanning windows. Afterward, we shuffled the order of these 5-msec windows over a 250-msec radius and then reconstructed them to create the scrambled speech sound. Meanwhile, we attempted to maintain the length, overall power, and frequency spectrum as similar as possible to those of the originally recorded phonemes. The AV stimuli and the duration of each auditory stimuli are presented in Table 2. The data of the control conditions were excluded from the later analysis.

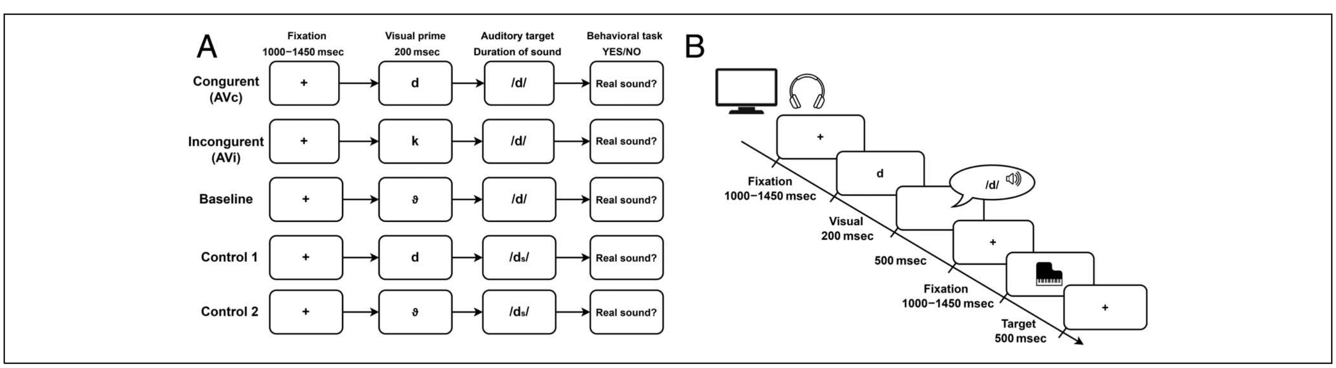

Figure 1. (A) Example trials for each condition of the behavioral priming task. /ds/ represents a scrambled speech sound. (B) The procedure used for the audiovisual priming paradigm in EEG data collection.

Table 2. Audiovisual Stimuli and Duration of Each Auditory Stimulus in the Priming Tasks

Letter EN		*Letter JP*		*Symbol*
d	171 msec	と (to)	327 msec	ϑ
k	143 msec	る (ru)	297 msec	ʉ
p	147 msec	ま (ma)	262 msec	ɫ
z	215 msec	ね (ne)	360 msec	ɕ

EN = English; JP = Japanese.

During the experiment, visual primes were followed by congruent and incongruent speech sounds. Participants were asked to attend to the visually presented kana/letter and then respond to the auditory stimuli. They needed to decide to use the "yes/no" buttons to determine whether each presented auditory stimulus was a "real sound" or not, as quickly as possible. The behavioral priming task was composed of four blocks for each language task, with the five mentioned conditions presented in a randomized sequence, totaling 48 trials for each condition, as illustrated in Figure 1A. All trials had a total duration of approximately 2000 msec. In each trial, a central fixation cross lasted ranging from 1000 to 1450 msec (150 msec/step) to minimize expectancy effects. Subsequently, a black-written visual prime was displayed on the white screen for a duration of 200 msec followed by the presence of the auditory stimuli. The kana/letters were presented in Arial font at a size of 28. The participants were instructed to prompt to provide a response after hearing the auditory stimuli. For each trial, we recorded both accuracy and RTs. Stimulus presentation and response collection were controlled using the E-Prime Version 3.0 software (Psychology Software Tools). Visual stimuli were displayed on a 21-in. screen with a screen resolution of 1920 × 1080 pixels at a distance of 60 cm, and auditory stimuli were binaurally presented by headphones from Panasonic. To familiarize themselves with the task, the participants were asked to perform practice trials until they achieved 80% accuracy. Each block lasted for approximately 2 min 30 sec. The entire experiment, including tasks for each language, took approximately 20 min. Participants had the opportunity to take a break after completing each block.

EEG Priming Measure

A slightly modified version of priming paradigms in Japanese and English was run using E-Prime (Version 2.0), while the EEG was being recorded. The duration of both visual and auditory stimuli aligned with the durations used in the behavioral experiments. One notable difference from the behavioral priming task was that we set a 500-msec duration as the interval from the onset of the speech sound to the presentation of a central fixation rather than the duration of sound itself to ensure a sufficiently long time window for the analysis of ERPs. Given that task-related top–down processing in an explicit matching task may impact the presence of the congruency effect in the auditory cortex (Mozolic, Hugenschmidt, Peiffer, & Laurienti, 2008), our task attempted to use a passive paradigm. As a second difference from behavioral priming task, we included one more condition (i.e., the target condition) to maintain participants' attention during the task. Consequently, this time, participants were not required to provide physical responses to the AV stimuli but needed to press the button when piano pictures were presented as targets (see Figure 1B). The participants sat in a sound-attenuating, electrically shielded room. All stimuli were presented on a 17-in. monitor with a screen resolution of 1024 × 720 pixels, and the participants sat 70 cm from the screen. The horizontal and vertical visual angles were 2.72° and 3.13°, respectively. The auditory stimuli were delivered binaurally through Panasonic earphones. Moreover, to enhance the reliability of the ERP data, the EEG experiment included nine blocks for each language, which made 108 trials for each condition, including 108 trials for the target condition. Within one block, all the conditions were presented in a randomized order. Participants initially completed tasks in either Japanese or English, followed by another language. The sequence of Japanese and English tasks was counterbalanced across participants. The entire experiment lasted approximately 50 min. The participants would have a rest upon their request.

EEG Recording and Preprocessing

Data were recorded using an electrode cap (Easycap GmbH) with 29 Ag/AgCl electrodes aligned according to the extended International 10–20 system (Fp1/2, F3/4, F7/8, Fz, FC1/2, FC5/6, T7/8, C3/4, Cz, CP5/6, CP1/2, P3/4, P7/8, Pz, POz, O1/2, and Oz). To monitor eye movements, the EOG was recorded simultaneously from an electrode placed 1 cm lateral to the outer canthus of each eye and from an electrode below the left eye (i.e., horizontal left and right EOG and vertical left EOG). All the channels were referenced to POz. EEGs and EOGs were amplified using a SynAmps amplifier (NeuroScan) at a sampling rate of 1000 Hz. Electrode impedance was maintained below 10 kΩ.

Preprocessing and offline data analysis were performed using the EEGLAB 14.1.2b toolbox (Delorme & Makeig, 2004), ERPLAB toolbox (Lopez-Calderon & Luck, 2014), and custom-made MATLAB scripts. The EEG signal from each electrode was rereferenced to the common average and then down-sampled to 250 Hz. For the data analysis, continuous data were high-pass filtered (noncausal Butterworth impulse response function, half-amplitude cutoff = 0.1 Hz, 12 dB/octave roll-off). EEG signals were then segmented into epochs from 600 msec before to

800 msec after the auditory onset. Independent component analysis was conducted to identify and remove components associated with eye movements, blink artifacts, and muscular activities. Components containing other nonbrain artifacts were identified using ICLabel (Pion-Tonachini, Kreutz-Delgado, & Makeig, 2019). Channels with larger artifacts were removed after visual inspection. The removed channels were interpolated using spherical interpolation. Afterward, epochs containing artifacts of amplitudes ±100 μV at the EOG and scalp electrodes were excluded from further analysis. We excluded the data of three participants from the ERP analyses (one AE, one IE from the English task, and one from the Japanese task, respectively), because of more than 30% of the trials rejected. The average number (*SD*) of epochs per group in each condition of the Japanese and English tasks is shown in Table 3. Before averaging for ERP, all epochs were filtered using a low-pass Butterworth filter (half-amplitude cutoff = 35 Hz, 12 dB/octave roll-off) and baseline-corrected using an interval from −300 to 0 msec before the onset of the sound.

Statistical Analysis

Statistical analyses were performed using SPSS Version 26.0 (SPSS, Inc.) and Stata Version 17.0 (StataCorp).

Behavioral Priming Measure Data

Accuracy (in percentage) and RTs (in milliseconds) in the Japanese and English tasks were computed for each participant in the behavioral priming task. Before analyzing RTs for responses, we excluded trials with values exceeding ± 2.5 *SD*s of an individual's average RT. A repeated-measures ANOVA was conducted to examine the RTs across the three conditions in the Japanese task. Similarly, we performed a repeated-measures ANOVA on the RTs obtained from the English task. The analysis incorporated factors Group (AE and IE) and Condition (congruent, incongruent, and baseline). To assess the normality assumptions, QQ plots were examined. The level of significance was set at a corrected $p < .05$. All reported p values from the repeated-measures ANOVA were Greenhouse–Geisser corrected. Bonferroni corrections were performed for all post hoc analyses. Reported values for the post hoc analyses were corrected.

EEG Priming Measure Data

The average of accuracy and RTs in the EEG priming measure for detecting targets were calculated to make sure of participants' engagement and compliance during the task. Statistical analyses of the auditory N1 and P2 components were carried out on epochs time-locked to the onset of auditory stimuli. The results of previous studies on typical topographical distribution have suggested AV congruency effects mostly observed at central sites (Knowland et al., 2014; Stekelenburg & Vroomen, 2007); thus, the apex, Cz, was selected for analyzing the N1 and P2 peak latencies and amplitudes. We employed a jackknife-based technique (Kiesel, Miller, Jolicoeur, & Brisson, 2008) to determine the peak latencies of the N1 and P2 components, to obtain more accurate and robust estimates of latency (Miller, Ulrich, & Schwarz, 2009). The jackknife procedure involves testing the latencies of N grand averages of $N - 1$ participants using a leave-one-out method. The N1 and P2 amplitudes were measured using the mean amplitudes over specific time windows centered on the mean peak latencies. We used the mean amplitude rather than the peak amplitude as it is generally regarded as a more reliable measurement. The N1 amplitude was chosen in a 110- to 140-msec time window after the onset of the auditory stimulus across all conditions and tasks. Similarly, the P2 amplitude was chosen in a 210- to 235-msec time window for both Japanese and English tasks. To assess the congruency effects, peak latencies and mean amplitudes were calculated for both the congruent and incongruent conditions. Specifically, we conducted pairwise comparisons of latency and amplitude between congruent and incongruent conditions in the Japanese task. In addition, repeated-measures ANOVAs were performed on the latency and amplitude measures of the English task, including one within-participant factor (congruent and incongruent conditions) and one between-participant factor (AE and IE groups). In addition, to investigate potential differences in spatially underlying processes, we conducted topographical analyses using a permutation test with 1000 randomizations in the time windows of N1 and P2, respectively. Furthermore, to compare our findings with those of Nash et al. (2017), latencies and amplitudes of N1 and P2 components for congruent and baseline trials were supplementarily analyzed. Finally, two-tailed Pearson's correlations were calculated to testify whether behavioral priming responses and mean amplitudes of electrodes

Table 3. Average Number (*SD*) of Epochs per Group in Each Condition of the Japanese and English Tasks

	Japanese Task (n = 32)	*English Task AE (n = 12)*	*English Task IE (n = 20)*
AVc	90.12 (±12.55)	86.08 (±12.88)	93.50 (±10.67)
AVi	92.03 (±11.72)	88.17 (±13.15)	94.70 (±10.66)
Baseline	90.72 (±13.57)	89.42 (±12.49)	94.95 (±11.20)

AVc = audiovisual congruent; AVi = audiovisual incongruent.

during the N1 time window were related to English language proficiency levels.

To ensure that the assumptions of normality were met, the QQ plots were inspected. Appropriate follow-up ANOVAs and pairwise comparisons were performed with corrections for multiple comparisons and an alpha level of .05. For topographical analysis, we adopted the false discovery rate method (Benjamini & Hochberg, 1995) for multiple comparisons.

RESULTS

Behavioral Priming Measure Results

Across conditions and groups, an average of 4.82 trials of RT data was excluded from the Japanese task, whereas an average of 6.30 trials of RT data was excluded from the English task. These excluded trials were not solely from the last block but rather distributed across multiple blocks. Hence, more than 97% of the possible RT data were used for the analysis, in each condition and for each group. The accuracy of the behavioral priming task was quite high for both the Japanese (mean = 97.54%, *SD* = 3.07%) and English (mean = 97.03%, *SD* = 2.93%) tasks, showing near-perfect levels of performance in the three conditions.

The mean RTs for both tasks are shown in Figure 2. For the Japanese task, faster RTs were observed in the congruent condition, in comparison to the incongruent or baseline conditions. Similarly, a congruency effect of the English task was found in both groups, with shorter RTs in the congruent condition than in the incongruent or baseline conditions.

A repeated-measures ANOVA of the three conditions in the Japanese task revealed a significant main effect of Condition, $F(2, 66) = 63.93, p < .001, \eta^2 = .66$. For RTs of the English task, the Condition × Group ANOVA yielded a significant main effect of Condition, $F(2, 64) = 38.92$, $p < .001$, $\eta^2 = .55$, but no significant effect of Group, $F(1, 32) = 1.97, p = .17, \eta^2 = .06$, nor a significant Group × Condition interaction, $F(2, 64) = 0.15, p = .82, \eta^2 = .01$. These findings revealed the presence of a congruency effect on behavioral responses, indicating the facilitation of congruent visual letters for the corresponding speech sound processing in both L1 and L2.

EEG Priming Measure Results

The EEG priming task yielded quite high accuracy rates for target detection, with an average of 97.02% for the Japanese task and 96.70% for the English task. Meanwhile, the mean RTs were 440.62 msec for the Japanese task and 433.17 msec for the English task, indicating consistent compliance among participants throughout the tasks.

Figure 3 shows the ERP waveforms at the Cz electrode in response to the auditory onset under the varying visual context conditions. In addition, the corresponding topographical maps within the N1 and P2 time windows are also presented in Figure 3 for the Japanese and English tasks in the two groups. Table 4 shows the grand-mean peak latency and mean amplitude values of the N1/P2 at the Cz electrode in the congruent and incongruent conditions for the Japanese task and English task of the two groups.

Japanese Task

A pairwise *t* test concerning peak latencies revealed that there was no difference between the congruent and incongruent conditions for N1 latency, $t(32) = -0.37, p = .72$, or P2 latency, $t(32) = -1.16, p = .25$, in the Japanese task. However, the pairwise comparison analysis of the N1 and P2 mean amplitudes demonstrated that the congruent condition differed significantly from the incongruent condition in terms of N1 amplitude, $t(32) = -2.06, p = .048$, and P2 amplitude, $t(32) = 2.05$, $p = .049$, respectively (Figure 4A).

For the topographical analysis, the results of the permutation tests between two conditions revealed one significant difference in each of the N1 and P2 time windows (Figure 3A.2). Specifically, within the N1 time window,

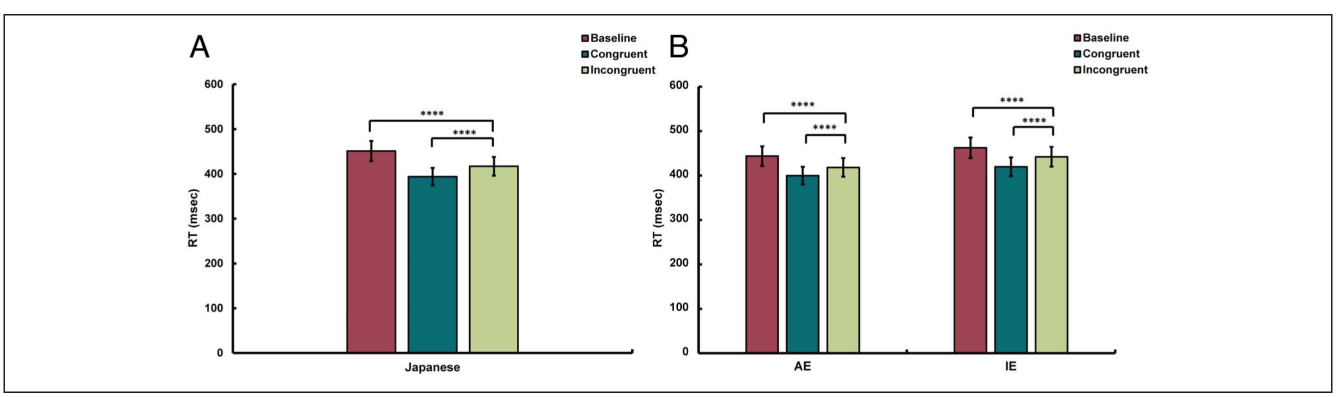

Figure 2. Mean RTs, with 95% within-participant confidence intervals, in the congruent, incongruent, and baseline conditions for the Japanese (A) and English (B) tasks for AE and IE learners. $^{****}p < .001$.

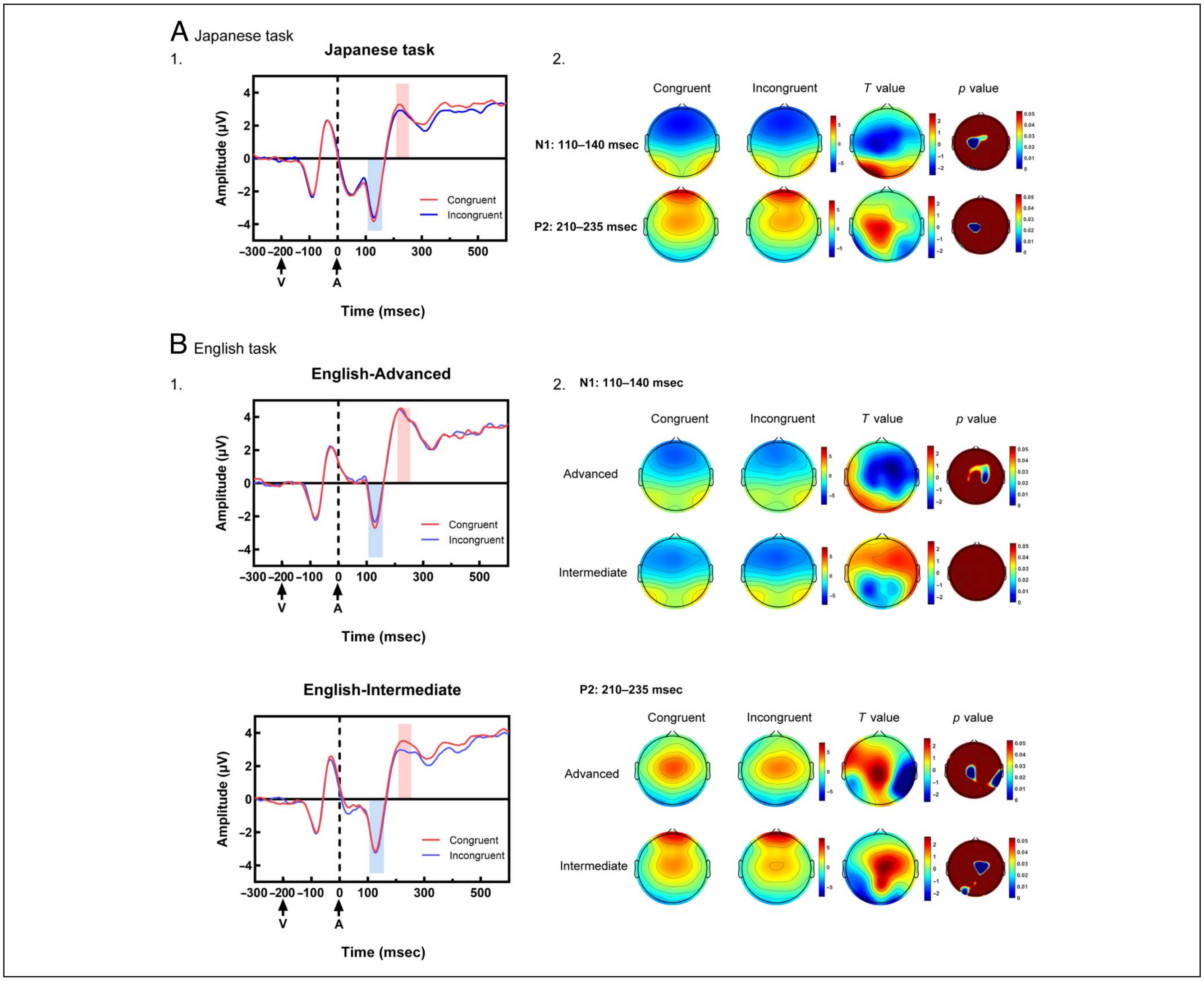

Figure 3. Averaged ERPs at the Cz electrode elicited by congruent trials (red) and incongruent trials (blue) in both the Japanese task (A.1) and the English task (B.1) for AE and IE learners. "V" (visual letter) and "A" (auditory stimulus) represent the onset of each stimulus. Topographical analyses illustrate averaged ERPs across electrodes in the N1 time window (110–140 msec) and the P2 time window (210–235 msec) for the Japanese (A.2) and English (B.2) tasks. The right column shows permutation-based p value distributions for statistical comparisons within conditions.

a significantly greater negative activity ($p < .05$) was observed at the central and left frontotemporal electrodes for congruent trials than for incongruent trials. In addition, another significant difference ($p < .05$) at the P2 time span was found in the left centrotemporal scalp regions, showing a greater amplitude for the congruent trials than for the incongruent trials.

Furthermore, to compare with our study findings with those of Nash et al. (2017), we compared the ERPs between congruent and baseline trials. The peak latencies were not significantly different between the congruent and baseline trials within the N1 time window, $t(32) = -0.98$, $p = .33$. However, a significant difference was identified in the P2 peak latency, $t(32) = -3.09$, $p = .004$. The follow-up analysis in the mean amplitudes of N1 and P2 components yielded a significant difference between congruent and baseline trials during the N1 time window, $t(32) = 2.32$, $p = .027$, but no significant difference was observed in the P2 component, $t(32) = 1.71$, $p = .097$.

English Task

For the English task, the repeated-measures ANOVA concerning N1 and P2 peak latencies revealed neither a significant main effect of Condition (N1: $F(1, 30) = 1.66$, $p = .21$, $\eta^2 = .05$; P2: $F(1, 30) = 1.23$, $p = .28$, $\eta^2 = .04$), nor a significant effect of Group (N1: $F(1, 30) = 0.49$, $p = .49$, $\eta^2 = .02$; P2: $F(1, 30) = 0.00$, $p = .96$, $\eta^2 = .00$). In addition, no significant Group × Condition interactions were observed for N1 latency, $F(1, 30) = 2.90$, $p = .099$, $\eta^2 = .09$, or P2 latency, $F(1, 30) = 0.49$, $p = .49$, $\eta^2 = .02$.

The Condition × Group ANOVA of the N1 mean amplitude (Figure 4B) showed neither a significant main effect

Table 4. Grand Means (*SE*s) of N1/P2 Peak Latency and Mean Amplitude at the Cz Electrode in the Two Conditions (AVc, AVi) for the Japanese Task and English Task of the Two Groups (AE and IE)

Task/Group	*ERP Component*	*Condition*	*Latency (msec)*	*Amplitude* (μ*V*)
JP	N1	AVc	128.85 (1.65)	−3.63 (0.44)
		AVi	129.33 (1.32)	−3.39 (0.42)
AE		AVc	130.67 (1.86)	−2.55 (0.67)
		AVi	130 (2.12)	−2.15 (0.62)
IE		AVc	126 (2.20)	−3.01 (0.59)
		AVi	130.8 (2.13)	−3.19 (0.60)
JP	P2	AVc	220.48 (2.62)	2.91 (0.35)
		AVi	223.27 (3.05)	2.59 (0.38)
AE		AVc	224 (4.43)	4.06 (0.36)
		AVi	223 (4.46)	3.98 (0.28)
IE		AVc	226 (4.41)	3.20 (0.54)
		AVi	221.6 (4.75)	2.58 (0.48)

AVc = audiovisual congruent; AVi = audiovisual incongruent; AE = advanced English; IE = intermediate English; JP = Japanese.

of Condition, $F(1, 30) = 0.73$, $p = .40$, $\eta^2 = .02$, nor a significant main effect of Group in the N1 component, $F(1, 30) = 0.68$, $p = .42$, $\eta^2 = .02$. The analysis of interaction revealed a significant interaction between Group and Condition on N1 amplitude, $F(1, 30) = 5.13$, $p = .031$, $\eta^2 = .15$. The subsequent post hoc pairwise comparisons did not reveal a significant difference between the congruent and incongruent conditions in either AE speakers, $t(11) = -1.70$, $p = .12$, or IE speakers, $t(19) = 1.29$, $p = .21$. Despite not reaching significance, the intermediate group demonstrated a tendency toward an incongruency effect, rather than a

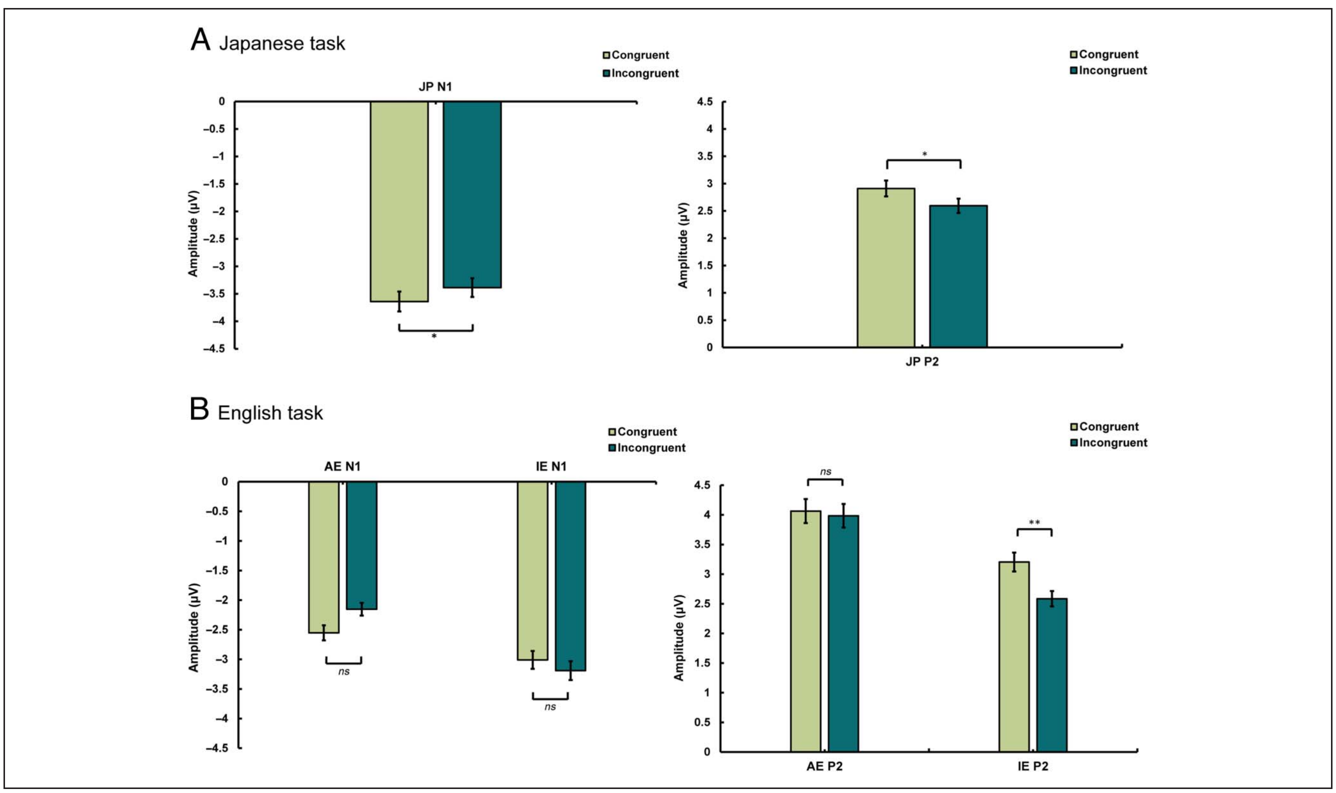

Figure 4. Grand mean amplitude values of N1 and P2 at the Cz electrode in congruent and incongruent conditions for the Japanese task (A) and the English task (B) for AE and IE learners. Error bars indicate 95% within-participant confidence intervals. *$p < .05$, **$p < .01$.

congruency effect. The results suggest the absence of congruent AV N1 enhancement regardless of the English proficiency level.

The analysis of the P2 amplitude (Figure 4B) revealed a significant main effect of Condition, $F(1, 30) = 4.98$, $p = .033$, $\eta^2 = .14$. However, we did not find a significant main effect of Group, $F(1, 30) = 2.69$, $p = .11$, $\eta^2 = .08$. In addition, the interaction analysis revealed no significant Group × Condition interaction in P2 amplitude, $F(1, 30) = 2.98$, $p = .095$, $\eta^2 = .09$. Exploratory post hoc pairwise comparisons showed that the conditions did not differ significantly among AE speakers, $t(11) = 0.34$, $p = .74$. However, the results showed a significant difference between the congruent and incongruent conditions for IE speakers, $t(19) = 3.14$, $p = .005$. As such, the results revealed the congruency effects among the IE speakers within the P2 time window.

The analysis of topographies revealed significant differences between the conditions among participants with different English proficiency levels (Figure 3B.2). For AE speakers, a larger N1 amplitude ($p < .05$) was found at the bilateral, yet predominantly right-lateralized, frontotemporal sensors for congruent trials compared to that for the incongruent trials. In addition, a significant difference ($p < .05$) within the P2 time window was found at the left temporal and central electrodes, with greater activity in the congruent trials than in the incongruent trials for AE speakers. In contrast, there was no significant difference at the N1 time span, but a greater amplitude for congruent trials was found at the right centrotemporal electrodes ($p < .05$) within the P2 time window in IE speakers.

As a supplementary analysis, we compared peak latencies and mean amplitudes between the congruent and baseline trials. Pairwise comparisons of peak latencies showed no significant differences between the congruent and baseline trials (N1: AE group, $t(11) = 0.32$, $p = .75$; IE group, $t(19) = -1.07$, $p = .30$; P2: AE group, $t(11) = 0.22$, $p = .83$; IE group, $t(19) = 1.19$, $p = .25$). Furthermore, analysis of the amplitudes in the N1 component revealed a significant difference between the congruent and baseline trials for the IE group, $t(19) = 3.18$, $p = .005$, but not for the AE group, $t(11) = 1.47$, $p = .17$. Meanwhile, the results of P2 amplitude analysis demonstrated that the congruent condition differed significantly from the baseline condition in both the AE group, $t(11) = 3.12$, $p = .010$, and the IE group, $t(19) = 3.15$, $p = .005$.

The Correlation between English Task and English Proficiency

Correlations were calculated to examine whether the English language proficiency level, as measured by L2 proficiency tests, was linked to the congruency effects in behavioral priming responses (i.e., RTs) and in ERP waveforms during the N1 time span. To gain a holistic understanding of the impact of language proficiency, the analyses were conducted without separating the participants into two groups. We employed the first principal component of the behavioral English proficiency tests as an index to represent the comprehensive English language proficiency level. On the basis of the results of topographical analysis, we chose the C3 and C4 electrodes, where the largest N1 congruency effects were observed. A small and nonsignificant correlation was found between L2 proficiency level and the behavioral index of congruency ($r = .06$, $p = .75$), suggesting that behavioral priming effects occurred regardless of participants' language proficiency. However, the L2 proficiency level correlated negatively with the difference of the mean amplitudes between the congruent and incongruent trials over the right (C4: $r = -.54$, $p = .001$; Figure 5) but not the left (C3: $r = -.32$, $p = .076$) centrotemporal electrodes. Therefore, a higher English proficiency level was significantly related to a stronger right-lateralized congruency effect in the bilateral frontotemporal regions. Furthermore, to investigate the potential association between English phonological and phonemic awareness and the N1 congruency effect, we therefore performed an additional correlation analysis between the second principal component of English proficiency tests and C3/C4 electrodes. Unfortunately, our analysis did not yield any significant correlation between phonological abilities and the N1 congruency effect on ERP amplitudes.

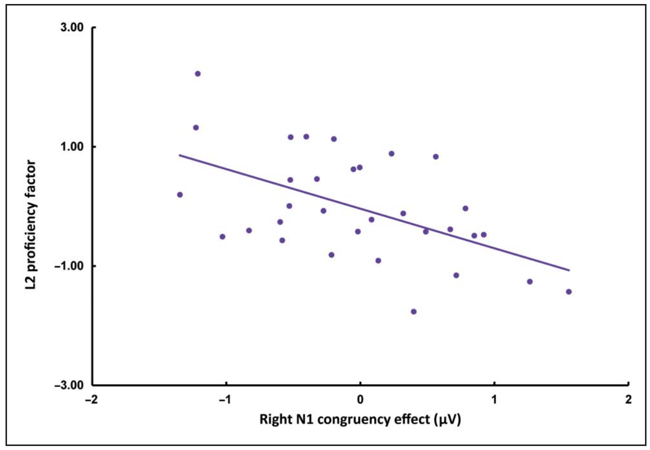

Figure 5. Correlations between the comprehensive English proficiency levels and congruency effects in the ERP data within N1 time window.

DISCUSSION

Using the behavioral performance and ERP responses, we investigated the role of L-SS integration in a nonalphabetic native language and its contribution to L2 reading in native Japanese adults with various English proficiency levels. In contrast to the previous study by Nash et al. (2017), we included an incongruent condition, enabling us to determine the validity of the congruency effect on automatic AV integration during language reading.

Behavioral Measure

The collected behavioral measures yielded significant results in RTs, showing an AV behavioral benefit in both Japanese and English tasks, with shorter RTs in the congruent condition than in the incongruent and baseline conditions. This finding tends to be consistent with previous findings (Clayton, West, Sears, Hulme, & Lervåg, 2020; Clayton & Hulme, 2018; Nash et al., 2017) on behavioral performance with the use of a priming paradigm in native English speakers. The following analyses of the English task showed no significant difference between groups in the relative size of the priming effect, irrespective of their English proficiency levels. In addition, we found only a small and nonsignificant correlation between the behavioral congruency effect and English proficiency levels, suggesting that congruency facilitation occurred regardless of participants' English proficiency level. In line with the lack of group differences in the English priming task, this result indicates that the behavioral priming task may not solely capture the processing of automatic L-SS integration.

ERP Measure

Regarding the effects of the visual context on auditory processing, previous ERP studies on AV speech integration have reported the shortening of auditory N1/P2 peak latency and the difference of amplitude in the presence of congruent AV stimuli, compared to that in the incongruent trials (e.g., Stekelenburg & Vroomen, 2007; van Wassenhove et al., 2005; Klucharev et al., 2003). In our study, the difference in the latency of the N1/P2 components, as shown in previous studies, was not found between the congruent and incongruent conditions for either the Japanese task or the English task. The findings of latency did not provide evidence of a speeding-up interpretation as in the behavioral results.

Contribution of L-SS Integration in Japanese Reading

For the Japanese task, the electrophysiological data revealed that the N1/P2 amplitudes elicited by the congruent condition were larger than those evoked by the incongruent condition. Our findings indicate that the congruent visual Japanese stimuli impacted the processing of auditory phonetic stimuli differentially. In addition, the larger N1 amplitude in response to congruent trials suggests that the physically congruent visual context serves as an informative cue that may facilitate the auditory processing by automatically activating the corresponding speech sound. A previous neuroimaging study (Herdman et al., 2006) in Japanese showed that visual congruency can modulate cortical activity in the left primary auditory cortex (i.e., left Heschl's sulcus/planum temporale) for typical adult Japanese readers. Similarly, fMRI studies indicated that successful integration of letters and speech sounds would lead to increased activities in the auditory cortex (e.g., Xu, Kolozsvári, Oostenveld, Leppänen, & Hämäläinen, 2019; Blau et al., 2008; van Atteveldt et al., 2004). Therefore, our findings on the scalp distribution of congruency effects observed in the left frontotemporal scalp region (within the N1 time window) could be explained by differential activities in the primary auditory cortex. The feedback hypothesis proposed by van Atteveldt et al. (2009) implies that the congruency effect in the auditory cortex is more likely to be caused by differential feedback from heteromodal areas in the STC. Hence, our findings suggest that the activation of the L-SS association may take place before the N1 time span, as the modulation in the auditory processing could only occur after the feedback sent from the STC. As an early N1 congruency effect was identified in the Japanese task, we confirmed the automatic activation of L-SS association by a visual letter in native Japanese speakers. These findings suggest that L-SS integration plays a similar role in Japanese reading as it does in alphabetic language reading.

Contribution of L-SS Integration in L2 Reading

The ERP data obtained from the English task did not reveal any significant congruency effects on the N1 mean amplitude in either the advanced group or the intermediate group. Nevertheless, the results of the topographical analysis demonstrated a bilateral frontotemporal congruency effect with a right-lateralized emphasis within the early auditory N1 time window, specifically in AE speakers. Accordingly, our data of the English task could be interpreted in a manner similar to that of the Japanese task. With increased L2 reading proficiency, the neural networks involved in the L-SS association in English become more efficient and can be rapidly activated, leading to the earlier emergence of congruency effects in the ERP waveform. Given the emergence of this right-lateralized congruency effect within the early N1 time window, we can infer that the visual English letters would activate the L-SS association in the STC, and subsequent feedback would be directed toward the primary auditory cortex, resulting in the observed congruency effects in the advanced group.

In contrast to our findings in the advanced group, recent fMRI findings in alphabetic languages have demonstrated greater left hemisphere and bilateral activation in response to congruent trials during AV integration (e.g., left: Pekkola et al., 2005; bilateral: Blau, van Atteveldt, Ekkebus, Goebel, & Blomert, 2009; Blau et al., 2008; van Atteveldt et al., 2004; also see a meta-analysis by Gao et al., 2023). Tervaniemi and Hugdahl (2003) proposed that the left and right hemispheres tend to be equally engaged in auditory language processing and suggested that relatively small changes in acoustic sound features or less familiarity could potentially influence the engagement levels of the left and right auditory areas in sound encoding. Moreover, Shtyrov, Pihko, and Pulvermüller (2005) suggested that the left-hemisphere lateralization of language is associated with the processing of large linguistic elements (e.g., words) rather than with the physical or phonological properties. In line with these perspectives, the bilateral but right-lateralized frontotemporal congruency effect observed in our study may be attributed to the use of English letters as stimuli in the task. In addition, studies on bilingualism have suggested that late L2 learners may show greater right hemispheric involvement in language processing, compared with that shown by monolinguals and early bilinguals (Cargnelutti, Tomasino, & Fabbro, 2019; Park, Badzakova-Trajkov, & Waldie, 2012; Dehaene et al., 1997). Given that our participants were late L2 learners, we can assume that their experience of learning English as an L2 may have potentially resulted in the enhanced right-lateralized neural activity in the advanced group. Another possible explanation could be that the right-lateralized congruency effect observed in our advanced group indeed reflects an "incomplete consolidation" as proposed by Eberhard-Moscicka, Jost, Raith, and Maurer (2015). They suggested a more right-lateralized N1 print tuning, observed as an increased visual N1 response to words during the early stages of reading training, possibly involving lexical–semantic processes. However, whether a similar pattern of "incomplete consolidation" might also be evident in the auditory cortex during L-SS integration needs further investigation before drawing stronger conclusions. Interestingly, we found that the right-lateralized early N1 congruency effect significantly correlated with the English proficiency level, suggesting its significant role in English reading fluency. Specifically, higher proficient readers demonstrated greater N1 amplitude in the congruent condition. Taken together, although the neural processing in advanced L2 learners may show some deviations from native language processing, this right-lateralized congruency effect within the N1 time window may be interpreted as the successful integration of English letters and sounds.

Regarding the intermediate group, we did not observe any significant congruency effect. This may indicate a failure of integrating English letters and sounds. Similar results have been reported in the previous MMN studies involving Chinese English learners (Du, Yang, Wang, Li, & Tao, 2023; Wang et al., 2019). Hence, our results suggest persistent difficulties in achieving effective L-SS integration during L2 reading among individuals with IE proficiency. Moreover, our results showed a tendency toward an N1 incongruency effect. Interestingly, despite variations in experimental designs and stimulus types, the direction of the effect (i.e., congruency/incongruency effects) is possibly related to reading proficiency. We noted that previous neuroimaging studies with typical adults have consistently reported increased activities in the auditory-related areas for the matching stimuli (e.g., Blau et al., 2008; van Atteveldt et al., 2004; Raij et al., 2000). In contrast, the reversed pattern, specifically the incongruency effect, tends to be observed in studies involving children with only a few years of formal reading instruction (Caffarra et al., 2021; Wang, Karipidis, Pleisch, Fraga-González, & Brem, 2020; Karipidis et al., 2017, 2018; Plewko et al., 2018). Therefore, we assume that this observed tendency toward an incongruent effect in our intermediate group was attributed to their English reading proficiency, suggesting an incomplete shift to fully automatic L-SS integration. Accordingly, it is plausible to consider that the achievement of the automaticity in L-SS integration for an L2 would require a longer learning time.

Furthermore, we found an enhancement of the P2 amplitude at Cz in response to the congruent condition compared with that in response to the incongruent condition among the intermediate group. Similarly, this congruency effect was evident in the Japanese task and was observed in the advanced proficiency group through topographical analysis. AV P2 suppression in the incongruent condition may be difficult to explain using the feedback hypothesis by van Atteveldt et al. (2009). However, there is much more consistency among studies on the auditory P2 modulations in the detecting and resolving conflicting information. For instance, Knowland et al. (2014) reported that P2 suppression of the incongruent condition in AV processing reflects the competition between conflicting multisensory inputs. Moreover, Crowley and Colrain (2004), in their review of the auditory P2 component, proposed that the P2 amplitude diminished as the attention increased. Similarly, the decreased P2 response elicited by incongruent AV stimuli in the intermediate group may be attributed to increased attentional resources for mismatched speech sounds. Furthermore, recent neuroimaging studies have shown that the right-lateralized auditory networks are associated with sustained attention and working memory (Hirnstein, Westerhausen, & Hugdahl, 2013; Corbetta & Shulman, 2002; Downar, Crawley, Mikulis, & Davis, 2001). To sum up, the observed congruency effects at the right centrotemporal electrodes might reflect increased auditory attention, rather than processes specifically related to the L-SS integration in IE learners. Thus, our findings provide evidence that attention may facilitate the successful activation of the L-SS association in the L2.

Comparison with the Study by Nash et al.

Indeed, it is important to note that the study by Nash et al. (2017) compared congruent AV pairs with the baseline condition, instead of the incongruent condition, which is different from previous fMRI studies (Blau et al., 2008; van Atteveldt et al., 2004, 2007). They assumed the baseline condition to be a relatively "neutral" condition, and the comparison between congruent and baseline conditions might be more suitable to measure the facilitation effects of congruent letter primes. In contrast, the incongruent trials tend to be involved with conflicting information, which may lead to slower responses. If this hypothesis holds true for adults, then we would expect shorter RTs in the baseline condition. However, our results revealed faster RTs in the incongruent condition rather than in the baseline condition. An alternative view is that these symbols may be perceived as technically "incongruent" with the speech sound, but this observed difference in RTs between incongruent and baseline conditions indicates different processing of the mismatch between visual prime and auditory stimuli for adults.

Moreover, Hein et al. (2007) suggested that unfamiliar visual stimuli and familiar but semantically incongruent stimuli might engage in different neural processing for adults. They found that unfamiliar pairs were integrated in the inferior frontal cortex, potentially reflecting the learning of novel AV associations. In light of their findings, we speculate that as the symbol stimuli in our study were less familiar to participants, they might have prompted to build a novel association between the visual symbol and the accompanying auditory stimulus. This additional involvement of the inferior frontal cortex could complicate the interpretation when comparing with the baseline condition. Taken together, although our study in comparison between congruent and baseline conditions found similar congruency effects of mean amplitudes in the time window of N1/P2 at the Cz electrode, as shown in the study by Nash et al. (2017), the results of the comparison between the congruent and baseline conditions in adult readers need to be interpreted with caution.

Limitations

There are some limitations of our study that should be taken into consideration. One limitation is the exclusive focus on L2 learners of English, without the inclusion of native English speakers. Although we attempted to compare our findings with those of Nash et al. (2017), it is important to note that their study involved English-speaking children rather than adults. To address this gap, future studies should include native adult English speakers to enable a more comprehensive comparison between children and adults. Furthermore, this study did not assess the Japanese proficiency of our participants. Although all participants were required to pass a rigorous examination in Japanese, our study did not specifically correlate their Japanese proficiency with the ERP data. Another limitation of this study is associated with the age of L2 acquisition and exposure to an L2. Our participants began to learn English after the entry to primary school, with a relatively late age of acquisition. Previous studies have shown that earlier exposure to an L2 may influence brain responses during L2 reading (Bowden et al., 2013; Cao et al., 2013). Hence, the results of this study may not be generalizable to individuals who begin learning English at an earlier age or have more extensive exposure to L2. Future studies could examine how the age of acquisition and exposure to the L2 affect L-SS integration among native Japanese learners. Moreover, we did not take into account L2s beyond English. Our participants primarily consisted of individuals proficient in native Japanese and using English as their L2, yet lacking the self-reported data on additional alphabetic or nonalphabetic languages. Proficiency in a third language, for instance, could potentially lead to different aspects of neurocognitive changes (e.g., Schroeder & Marian, 2017). Therefore, future studies will need to consider not only task-related factors but also individual differences in language experience for a more comprehensive analysis.

To conclude, the present study explored the role of L-SS integration during native Japanese and L2 reading among native Japanese speakers with varying levels of English proficiency. The main findings of our study can be summarized as follows. In the behavioral experiment, we observed shorter RTs in the congruent condition than in the incongruent and baseline conditions for both the Japanese and English tasks, suggesting a facilitation effect of congruency, regardless of language proficiency. The ERP experiment results, on the other hand, demonstrated the congruency effect of the early auditory N1 response in the left frontotemporal scalp region during native Japanese reading. Interestingly, increased activity in the bilateral but predominantly right-lateralized frontotemporal areas elicited by the congruent condition within the N1 time window was found in AE learners, and this right-lateralized early congruency effect was significantly correlated with English ability. Our results for intermediate learners showed a right-lateralized congruency effect on P2, suggesting that increased attentional resources might be needed. On the basis of our findings, we therefore conjecture that despite some deviations from L1 processing, AE learners may successfully integrate letters and sounds in English reading, whereas intermediate learners might encounter difficulty in achieving L-SS integration in L2 reading in contrast to their native language. Besides, L2 proficiency may have an impact on the level of automaticity in L-SS integration, and in support of a compensatory role, the right P2 congruency effect may be necessary for nonnative language learners to aid lower levels of automaticity. Our study can shed light on fundamental reading mechanisms in the nonalphabetic language as well as in the alphabetic L2. These findings can serve as a foundation for research investigating underlying neuronal abnormalities among individuals with difficulties in L1 and L2 reading.

APPENDIX

Table A1. Summary of Principal Component Analysis: Principal Component Loadings of Each Variable Contributing to Two Identified Components

	Principal Component	
Variables	*1*	*2*
TOWRE II Sight Word Efficiency	**.78**	.20
TOWRE II Phonemic Decoding Efficiency	**.67**	**.54**
WJ III LWI	**.71**	.29
WJ III WA	**.77**	.35
WJ III RF	**.91**	.11
WJ III SS	**.74**	.19
CTOPP-2 Elision	.13	**.81**
CTOPP-2 Blending Words	.48	**.62**
CTOPP-2 Blending Nonwords	**.66**	.44
CTOPP-2 Segmenting Nonwords	.22	**.84**

Boldface indicates the testing measurement contributing to the principal component above .5, which is deemed with stronger correlation with the component. TOWRE II Phonemic Decoding Efficiency shows a stronger correlation with the two identified components, suggesting its contribution to both comprehensive reading ability and phonemic/phonological ability. TOWRE = Test of Word Reading Efficiency; WJ = Woodcock–Johnson; LWI = Letter–Word Identification; WA = Word Attack; RF = Reading Fluency; SS = Spelling of Sounds; CTOPP-2 = Comprehensive Test of Phonological Processing–Second Edition.

Acknowledgments

This work was supported by Japan Society for the Promotion of Science, JSPS KAKENHI, Grant Number JP17H02713.

Corresponding author: Dongyang Yan, Graduate School of Education, Hokkaido University, Nishi-7, Kita-11, Kita-ku, Sapporo, Hokkaido, Japan, or via e-mail: ydyxj@hotmail.com.

Data Availability Statement

The data sets generated and/or analyzed during the current study are available from the corresponding author on reasonable request.

Author Contributions

Dongyang Yan: Conceptualization; Data curation; Formal analysis; Investigation; Visualization; Writing—Original draft; Writing—Review & editing. Ayumi Seki: Conceptualization; Funding acquisition; Resources; Supervision; Writing—Review & editing.

Diversity in Citation Practices

Retrospective analysis of the citations in every article published in this journal from 2010 to 2021 reveals a persistent pattern of gender imbalance: Although the proportions of authorship teams (categorized by estimated gender identification of first author/last author) publishing in the *Journal of Cognitive Neuroscience (JoCN)* during this period were M(an)/M = .407, W(oman)/M = .32, M/W = .115, and W/W = .159, the comparable proportions for the articles that these authorship teams cited were M/M = .549, W/M = .257, M/W = .109, and W/W = .085 (Postle and Fulvio, *JoCN*, 34:1, pp. 1–3). Consequently, *JoCN* encourages all authors to consider gender balance explicitly when selecting which articles to cite and gives them the opportunity to report their article's gender citation balance.

REFERENCES

Andres, A. J. D., Oram Cardy, J. E. O., & Joanisse, M. F. (2011). Congruency of auditory sounds and visual letters modulates mismatch negativity and P300 event-related potentials. *International Journal of Psychophysiology*, *79*, 137–146. https://doi.org/10.1016/j.ijpsycho.2010.09.012, PubMed: 20888870

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*, 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Bernstein, L. E., Auer, E. T., Jr., Wagner, M., & Ponton, C. W. (2008). Spatiotemporal dynamics of audiovisual speech processing. *Neuroimage*, *39*, 423–435. https://doi.org/10.1016/j.neuroimage.2007.08.035, PubMed: 17920933

Blau, V., van Atteveldt, N., Ekkebus, M., Goebel, R., & Blomert, L. (2009). Reduced neural integration of letters and speech sounds links phonological and reading deficits in adult dyslexia. *Current Biology*, *19*, 503–508. https://doi.org/10.1016/j.cub.2009.01.065, PubMed: 19285401

Blau, V., van Atteveldt, N., Formisano, E., Goebel, R., & Blomert, L. (2008). Task-irrelevant visual letters interact with the processing of speech sounds in heteromodal and unimodal cortex. *European Journal of Neuroscience*, *28*, 500–509. https://doi.org/10.1111/j.1460-9568.2008.06350.x, PubMed: 18702722

Boersma, P., & Weenink, D. (2020). *Praat: Doing phonetics by computer* (Version 4.1) [Computer program]. Retrieved June 6, 2020, from www.praat.org.

Bowden, H. W., Steinhauer, K., Sanz, C., & Ullman, M. T. (2013). Native-like brain processing of syntax can be attained by university foreign language learners. *Neuropsychologia*, *51*, 2492–2511. https://doi.org/10.1016/j.neuropsychologia.2013.09.004, PubMed: 24051003

Caffarra, S., Lizarazu, M., Molinaro, N., & Carreiras, M. (2021). Reading-related brain changes in audiovisual processing: Cross-sectional and longitudinal MEG evidence. *Journal of Neuroscience*, *41*, 5867–5875. https://doi.org/10.1523/JNEUROSCI.3021-20.2021, PubMed: 34088796

Cao, F., Tao, R., Liu, L., Perfetti, C. A., & Booth, J. R. (2013). High proficiency in a second language is characterized by greater involvement of the first language network: Evidence from Chinese learners of English. *Journal of Cognitive Neuroscience*, *25*, 1649–1663. https://doi.org/10.1162/jocn_a_00414, PubMed: 23654223

Cargnelutti, E., Tomasino, B., & Fabbro, F. (2019). Language brain representation in bilinguals with different age of appropriation and proficiency of the second language: A meta-analysis of functional imaging studies. *Frontiers in Human Neuroscience*, *13*, 154. https://doi.org/10.3389/fnhum.2019.00154, PubMed: 31178707

Clayton, F. J., & Hulme, C. (2018). Automatic activation of sounds by letters occurs early in development but is not impaired in children with dyslexia. *Scientific Studies of Reading*, *22*, 137–151. https://doi.org/10.1080/10888438.2017.1390754

Clayton, F. J., West, G., Sears, C., Hulme, C., & Lervåg, A. (2020). A longitudinal study of early reading development: Letter–sound knowledge, phoneme awareness and RAN, but not letter–sound integration, predict variations in reading development. *Scientific Studies of Reading*, *24*, 91–107. https://doi.org/10.1080/10888438.2019.1622546

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience*, *3*, 201–215. https://doi.org/10.1038/nrn755, PubMed: 11994752

Crowley, K. E., & Colrain, I. M. (2004). A review of the evidence for P2 being an independent component process: Age, sleep and modality. *Clinical Neurophysiology*, *115*, 732–744. https://doi.org/10.1016/j.clinph.2003.11.021, PubMed: 15003751

Dehaene, S., Dupoux, E., Mehler, J., Cohen, L., Paulesu, E., Perani, D., et al. (1997). Anatomical variability in the cortical representation of first and second language. *NeuroReport*, *8*, 3809–3815. https://doi.org/10.1097/00001756-199712010-00030, PubMed: 9427375

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*, 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009, PubMed: 15102499

Dickens, R. H., Meisinger, E. B., & Tarar, J. M. (2015). Comprehensive Test of Phonological Processing (CTOPP-2). *Canadian Journal of School Psychology*, *30*, 155–162. https://doi.org/10.1177/0829573514563280

Dijkstra, T., Schreuder, R., & Frauenfelder, U. H. (1989). Grapheme context effects on phonemic processing. *Language and Speech*, *32*, 89–108. https://doi.org/10.1177/002383098903200201

Downar, J., Crawley, A. P., Mikulis, D. J., & Davis, K. D. (2001). The effect of task relevance on the cortical response to changes in visual and auditory stimuli: An event-related fMRI study. *Neuroimage*, *14*, 1256–1267. https://doi.org/10.1006/nimg.2001.0946, PubMed: 11707082

Du, B., 杜彬, Yang, Z., 杨振, Wang, C., 王翠, Li, Y., 李媛, & Tao, S., 陶沙. (2023). Short-term training helps second-language learners read like native readers: An ERP study. *Brain and Language*, *239*, 105251. https://doi.org/10.1016/j.bandl.2023.105251, PubMed: 36931112

Eberhard-Moscicka, A. K., Jost, L. B., Daum, M. M., & Maurer, U. (2021). Predicting reading from behavioral and neural measures—A longitudinal event-related potential study. *Frontiers in Psychology*, *12*, 733494. https://doi.org/10.3389/fpsyg.2021.733494, PubMed: 34916991

Eberhard-Moscicka, A. K., Jost, L. B., Raith, M., & Maurer, U. (2015). Neurocognitive mechanisms of learning to read: Print tuning in beginning readers related to word-reading fluency and semantics but not phonology. *Developmental Science*, *18*, 106–118. https://doi.org/10.1111/desc.12189, PubMed: 24863157

Ehri, L. C. (2005). Learning to read words: Theory, findings, and issues. *Scientific Studies of Reading*, *9*, 167–188. https://doi.org/10.1207/s1532799xssr0902_4

Ellis, D. P. W. (2010). Time-domain scrambling of audio signals in MATLAB. Retrieved from www.ee.columbia.edu/~dpwe/resources/matlab/scramble/.

Fitzgerald, K., & Todd, J. (2020). Making sense of mismatch negativity. *Frontiers in Psychiatry*, *11*, 468. https://doi.org/10.3389/fpsyt.2020.00468, PubMed: 32595529

Froyen, D. J. W., Bonte, M. L., van Atteveldt, N., & Blomert, L. (2009). The long road to automation: Neurocognitive development of letter–speech sound processing. *Journal of Cognitive Neuroscience*, *21*, 567–580. https://doi.org/10.1162/jocn.2009.21061, PubMed: 18593266

Froyen, D., Van Atteveldt, N., Bonte, M., & Blomert, L. (2008). Cross-modal enhancement of the MMN to speech-sounds indicates early and automatic integration of letters and speech-sounds. *Neuroscience Letters*, *430*, 23–28. https://doi.org/10.1016/j.neulet.2007.10.014, PubMed: 18023979

Froyen, D., Willems, G., & Blomert, L. (2011). Evidence for a specific cross-modal association deficit in dyslexia: An electrophysiological study of letter–speech sound processing. *Developmental Science*, *14*, 635–648. https://doi.org/10.1111/j.1467-7687.2010.01007.x, PubMed: 21676085

Gao, C., Green, J. J., Yang, X., Oh, S., Kim, J., & Shinkareva, S. V. (2023). Audiovisual integration in the human brain: A coordinate-based meta-analysis. *Cerebral Cortex*, *33*, 5574–5584. https://doi.org/10.1093/cercor/bhac443, PubMed: 36336347

Grill-Spector, K., & Malach, R. (2001). fMR-adaptation: A tool for studying the functional properties of human cortical neurons. *Acta Psychologica*, *107*, 293–321. https://doi.org/10.1016/S0001-6918(01)00019-1, PubMed: 11388140

Hein, G., Doehrmann, O., Müller, N. G., Kaiser, J., Muckli, L., & Naumer, M. J. (2007). Object familiarity and semantic congruency modulate responses in cortical audiovisual integration areas. *Journal of Neuroscience*, *27*, 7881–7887. https://doi.org/10.1523/JNEUROSCI.1740-07.2007, PubMed: 17652579

Henson, R. N. A. (2003). Neuroimaging studies of priming. *Progress in Neurobiology*, *70*, 53–81. https://doi.org/10.1016/S0301-0082(03)00086-8, PubMed: 12927334

Herdman, A. T., Fujioka, T., Chau, W., Ross, B., Pantev, C., & Picton, T. W. (2006). Cortical oscillations related to processing congruent and incongruent grapheme–phoneme pairs. *Neuroscience Letters*, *399*, 61–66. https://doi.org/10.1016/j.neulet.2006.01.069, PubMed: 16507333

Hirnstein, M., Westerhausen, R., & Hugdahl, K. (2013). The right planum temporale is involved in stimulus-driven, auditory attention—Evidence from transcranial magnetic stimulation. *PLoS One*, *8*, e57316. https://doi.org/10.1371/journal.pone.0057316, PubMed: 23437367

Hoonhorst, I., Serniclaes, W., Collet, G., Colin, C., Markessis, E., Radeau, M., et al. (2009). N1b and Na subcomponents of the N100 long latency auditory evoked-potential: Neurophysiological correlates of voicing in French-speaking subjects. *Clinical Neurophysiology*, *120*, 897–903. https://doi.org/10.1016/j.clinph.2009.02.174, PubMed: 19329357

Jost, L. B., Eberhard-Moscicka, A. K., Frisch, C., Dellwo, V., & Maurer, U. (2014). Integration of spoken and written words in beginning readers: A topographic ERP study. *Brain Topography*, *27*, 786–800. https://doi.org/10.1007/s10548-013-0336-4, PubMed: 24271979

Kaganovich, N., & Schumaker, J. (2014). Audiovisual integration for speech during mid-childhood: Electrophysiological evidence. *Brain and Language*, *139*, 36–48. https://doi.org/10.1016/j.bandl.2014.09.011, PubMed: 25463815

Karipidis, I. I., Pleisch, G., Brandeis, D., Roth, A., Röthlisberger, M., Schneebeli, M., et al. (2018). Simulating reading acquisition: The link between reading outcome and multimodal brain signatures of letter–speech sound learning in prereaders. *Scientific Reports*, *8*, 7121. https://doi.org/10.1038/s41598-018-24909-8, PubMed: 29740067

Karipidis, I. I., Pleisch, G., Röthlisberger, M., Hofstetter, C., Dornbierer, D., Stämpfli, P., et al. (2017). Neural initialization of audiovisual integration in prereaders at varying risk for developmental dyslexia. *Human Brain Mapping*, *38*, 1038–1055. https://doi.org/10.1002/hbm.23437, PubMed: 27739608

Kiesel, A., Miller, J., Jolicoeur, P., & Brisson, B. (2008). Measurement of ERP latency differences: A comparison of single-participant and jackknife-based scoring methods. *Psychophysiology*, *45*, 250–274. https://doi.org/10.1111/j.1469-8986.2007.00618.x, PubMed: 17995913

Klucharev, V., Möttönen, R., & Sams, M. (2003). Electrophysiological indicators of phonetic and non-phonetic multisensory interactions during audiovisual speech perception. *Cognitive Brain Research*, *18*, 65–75. https://doi.org/10.1016/j.cogbrainres.2003.09.004, PubMed: 14659498

Knowland, V. C. P., Mercure, E., Karmiloff-Smith, A., Dick, F., & Thomas, M. S. C. (2014). Audio-visual speech perception: A developmental ERP investigation. *Developmental Science*, *17*, 110–124. https://doi.org/10.1111/desc.12098, PubMed: 24176002

Li, P., Zhang, F., Yu, A., & Zhao, X. (2020). Language History Questionnaire (LHQ3): An enhanced tool for assessing multilingual experience. *Bilingualism: Language and Cognition*, *23*, 938–944. https://doi.org/10.1017/S1366728918001153

Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: An open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience*, *8*, 213. https://doi.org/10.3389/fnhum.2014.00213, PubMed: 24782741

Miller, J., Ulrich, R., & Schwarz, W. (2009). Why jackknifing yields good latency estimates. *Psychophysiology*, *46*, 300–312. https://doi.org/10.1111/j.1469-8986.2008.00761.x, PubMed: 19207197

Mozolic, J. L., Hugenschmidt, C. E., Peiffer, A. M., & Laurienti, P. J. (2008). Modality-specific selective attention attenuates multisensory integration. *Experimental Brain Research*, *184*, 39–52. https://doi.org/10.1007/s00221-007-1080-3, PubMed: 17684735

Murray, M. M., & Wallace, M. T. (2011). *The neural bases of multisensory processes* (1st ed.). Boca Raton, FL: CRC Press. https://doi.org/10.1201/9781439812174

Nakada, T., Fujii, Y., & Kwee, I. L. (2001). Brain strategies for reading in the second language are determined by the first language. *Neuroscience Research*, *40*, 351–358. https://doi.org/10.1016/S0168-0102(01)00247-4, PubMed: 11463481

Nash, H. M., Gooch, D., Hulme, C., Mahajan, Y., McArthur, G., Steinmetzger, K., et al. (2017). Are the literacy difficulties that characterize developmental dyslexia associated with a failure to integrate letters and speech sounds? *Developmental Science*, *20*, e12423. https://doi.org/10.1111/desc.12423, PubMed: 27496263

Pang, E. W., & Taylor, M. J. (2000). Tracking the development of the N1 from age 3 to adulthood: An examination of speech and non-speech stimuli. *Clinical Neurophysiology*, *111*, 388–397. https://doi.org/10.1016/S1388-2457(99)00259-X, PubMed: 10699397

Park, H. R. P., Badzakova-Trajkov, G., & Waldie, K. E. (2012). Language lateralisation in late proficient bilinguals: A lexical decision fMRI study. *Neuropsychologia*, *50*, 688–695. https://doi.org/10.1016/j.neuropsychologia.2012.01.005, PubMed: 22245007

Pekkola, J., Ojanen, V., Autti, T., Jääskeläinen, I. P., Möttönen, R., Tarkiainen, A., et al. (2005). Primary auditory cortex activation by visual speech: An fMRI study at 3 T. *NeuroReport*, *16*, 125–128. https://doi.org/10.1097/00001756-200502080-00010, PubMed: 15671860

Perani, D., & Abutalebi, J. (2005). The neural basis of first and second language processing. *Current Opinion in Neurobiology*, *15*, 202–206. https://doi.org/10.1016/j.conb.2005.03.007, PubMed: 15831403

Pion-Tonachini, L., Kreutz-Delgado, K., & Makeig, S. (2019). ICLabel: An automated electroencephalographic independent component classifier, dataset, and website. *Neuroimage*, *198*, 181–197. https://doi.org/10.1016/j.neuroimage.2019.05.026, PubMed: 31103785

Plewko, J., Chyl, K., Bola, Ł., Łuniewska, M., Dębska, A., Banaszkiewicz, A., et al. (2018). Letter and speech sound association in emerging readers with familial risk of dyslexia. *Frontiers in Human Neuroscience*, *12*, 393. https://doi.org/10.3389/fnhum.2018.00393, PubMed: 30333739

Raij, T., Uutela, K., & Hari, R. (2000). Audiovisual integration of letters in the human brain. *Neuron*, *28*, 617–625. https://doi.org/10.1016/S0896-6273(00)00138-0, PubMed: 11144369

Roa Romero, Y., Senkowski, D., & Keil, J. (2015). Early and late beta-band power reflect audiovisual perception in the McGurk illusion. *Journal of Neurophysiology*, *113*, 2342–2350. https://doi.org/10.1152/jn.00783.2014, PubMed: 25568160

Schrank, F. A. (2010). Woodcock–Johnson III tests of cognitive abilities. In A. S. Davis (Ed.), *Handbook of pediatric neuropsychology* (pp. 415–434). New York: Springer.

Schroeder, S. R., & Marian, V. (2017). Cognitive consequences of trilingualism. *International Journal of Bilingualism*, *21*, 754–773. https://doi.org/10.1177/1367006916637288, PubMed: 29151803

Shtyrov, Y., Pihko, E., & Pulvermüller, F. (2005). Determinants of dominance: Is language laterality explained by physical or linguistic features of speech? *Neuroimage*, *27*, 37–47. https://doi.org/10.1016/j.neuroimage.2005.02.003, PubMed: 16023039

Stekelenburg, J. J., & Vroomen, J. (2007). Neural correlates of multisensory integration of ecologically valid audiovisual events. *Journal of Cognitive Neuroscience*, *19*, 1964–1973. https://doi.org/10.1162/jocn.2007.19.12.1964, PubMed: 17892381

Tervaniemi, M., & Hugdahl, K. (2003). Lateralization of auditory-cortex functions. *Brain Research Reviews*, *43*, 231–246. https://doi.org/10.1016/j.brainresrev.2003.08.004, PubMed: 14629926

Torgeson, J. K., Wagner, R. K., & Rashotte, C. A. (1999). *Test review: Test of Word Reading Efficiency (TOWRE)*. Austin, TX: Pro-Ed.

van Atteveldt, N., Formisano, E., Goebel, R., & Blomert, L. (2004). Integration of letters and speech sounds in the human brain. *Neuron*, *43*, 271–282. https://doi.org/10.1016/j.neuron.2004.06.025, PubMed: 15260962

van Atteveldt, N. M., Formisano, E., Blomert, L., & Goebel, R. (2007). The effect of temporal asynchrony on the multisensory integration of letters and speech sounds. *Cerebral Cortex*, *17*, 962–974. https://doi.org/10.1093/cercor/bhl007, PubMed: 16751298

van Atteveldt, N. M., Formisano, E., Goebel, R., & Blomert, L. (2007). Top–down task effects overrule automatic multisensory responses to letter–sound pairs in auditory association cortex. *Neuroimage*, *36*, 1345–1360. https://doi.org/10.1016/j.neuroimage.2007.03.065, PubMed: 17513133

van Wassenhove, V., Grant, K. W., & Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proceedings of the National Academy of Sciences, U.S.A.*, *102*, 1181–1186. https://doi.org/10.1073/pnas.0408949102, PubMed: 15647358

van Atteveldt, N., Roebroeck, A., & Goebel, R. (2009). Interaction of speech and script in human auditory cortex: Insights from neuro-imaging and effective connectivity. *Hearing Research*, *258*, 152–164. https://doi.org/10.1016/j.heares.2009.05.007, PubMed: 19500658

Wallace, M. T., & Stevenson, R. A. (2014). The construct of the multisensory temporal binding window and its dysregulation in developmental disabilities. *Neuropsychologia*, *64*, 105–123. https://doi.org/10.1016/j.neuropsychologia.2014.08.005, PubMed: 25128432

Wang, C., Flemming, K., Yang, Z., Cortiana, G., Lammert, J., Rafat, Y., et al. (2022). Second language immersion experience could help the brain response to second language reading for native Chinese speakers. *Journal of Cognitive Neuroscience*, *34*, 2311–2319. https://doi.org/10.1162/jocn_a_01913, PubMed: 36122357

Wang, F., Karipidis, I. I., Pleisch, G., Fraga-González, G., & Brem, S. (2020). Development of print-speech integration in the brain of beginning readers with varying reading skills. *Frontiers in Human Neuroscience*, *14*, 289. https://doi.org/10.3389/fnhum.2020.00289, PubMed: 32922271

Wang, C., Yang, Z., Cao, F., Liu, L., & Tao, S. (2019). Letter–sound integration in native Chinese speakers learning English: Brain fails in automatic responses but succeeds with more attention. *Cognitive Neuroscience*, *10*, 100–116. https://doi.org/10.1080/17588928.2018.1529665, PubMed: 30270811

Wechsler, D. (1999). *Wechsler abbreviated scale of intelligence*. San Antonio, TX: The Psychological Corporation, Harcourt Brace Company. https://doi.org/10.1037/t15170-000

Xu, W., Kolozsvári, O. B., Oostenveld, R., Leppänen, P. H. T., & Hämäläinen, J. A. (2019). Audiovisual processing of Chinese characters elicits suppression and congruency effects in MEG. *Frontiers in Human Neuroscience*, *13*, 18. https://doi.org/10.3389/fnhum.2019.00018, PubMed: 30787872

Žarić, G., González, G. F., Tijms, J., van der Molen, M. W., Blomert, L., & Bonte, M. (2014). Reduced neural integration of letters and speech sounds in dyslexic children scales with individual differences in reading fluency. *PLoS One*, *9*, e110337. https://doi.org/10.1371/journal.pone.0110337, PubMed: 25329388

Zhou, H.-Y., Cheung, E. F. C., & Chan, R. C. K. (2020). Audiovisual temporal integration: Cognitive processing, neural mechanisms, developmental trajectory and potential interventions. *Neuropsychologia*, *140*, 107396. https://doi.org/10.1016/j.neuropsychologia.2020.107396, PubMed: 32087206