Cross-modal priming of written words across different timing conditions

Johanna Funk*[1,2], Louis T. Hessler Carbonell[1], Denise Al-Rubaye-Jung[1], Michelle Morgenstern[1], Elisa Schmied[1], Jasemina Tzallas[1], and Florian Hintz[1,2,3]

[1]Philipps University of Marburg, Marburg [2]Center for Mind, Brain and Behavior, Philipps-Universität Marburg and Justus-Liebig-Universität Giessen [3]Max Planck Institute for Psycholinguistics, Nijmegen
*corresponding author – johanna.funk@uni-marburg.de

Abstract

The present study investigated the effects of presentational timing, operationalized as different levels of temporal overlap, on cross-modal priming of written words. We used a paradigm where the playback of spoken word primes was shifted relative to the presentation of written targets (asynchronous, partially overlapping, and synchronous presentation). Our participants (n = 48) carried out a speeded lexical decision task on the written targets. Presenting the spoken primes, albeit the words' onset, before the written targets reduced lexical decision times to both words and pseudowords. Asynchronous presentation of the spoken primes resulted in the largest difference between word and pseudoword response times. We discuss our results in relation to the mental structure of human word knowledge and in the context of word form acquisition.

Keywords: word recognition; priming; presentational timing

Introduction

The acquisition of literacy fundamentally changes how word knowledge is stored in the human mind and brain (Ziegler & Goswami, 2005, for review). One hallmark is that the mental representation of spoken words is augmented with orthographic information. Beginning readers associate the phonemes that make up spoken words with the corresponding graphemes of the written word forms. As readers become more proficient, the associations between phonemes and graphemes become more entrenched, and the mutual activation of phonological and orthographic representations turns into an automatic process. That is, there is growing evidence that upon recognizing a spoken word its orthographic code is activated as well (Perre et al., 2009; Rastle et al., 2011). Similarly, there is a large body of experimental and computational research demonstrating that recognizing written words leads to the retrieval of the corresponding phonological forms (Frost, 1998; Rastle & Brysbaert, 2006). Phonological recoding, as the process is often referred to (McCusker et al., 1981), has turned into an integral part of models and theories of reading (Diependaele et al., 2010; Harm & Seidenberg, 2004; Perfetti, 2007). In fact, Perfetti (2007) argues that efficient reading comprehension hinges on the *synchronous* retrieval of written and spoken word forms—a claim that is supported by experimental work in dyslexic individuals who showed evidence for asynchronous (i.e., delayed) retrieval of phonological forms during reading (Breznitz & Misra, 2003). Perfetti's (2007) claim also fits well with experimental evidence that literacy programs that focus on orthography-phonology relationships ('phonics approach') are generally more efficient for literacy acquisition than programs focusing on orthography-semantics relationships (Taylor et al., 2017). Taylor and colleagues (2017) and other advocates of the 'phonics approach' thus stress the importance of a tight coupling and coordinated interaction of mental orthographic and phonological representations.

Although the co-activation of orthographic and phonologic information during spoken and written word processing has received plenty of experimental support, most models of word recognition focus on one of the modalities (spoken: Marslen-Wilson, 1987; McClelland & Elman, 1986; written: Coltheart et al., 2001). For example, the Cohort model of spoken word recognition posits that potential target candidates are activated in a bottom-up manner by the incoming speech signal. This cohort is continuously updated as new phonological input comes in, until one candidate remains (e.g., Marslen-Wilson, 1987).

A model that is concerned with the co-activation of phonological and orthographic codes during word processing, is the Bimodal Interactive Activation Model (BIAM; (Diependaele et al., 2010; Grainger et al., 2003; Grainger & Ferrand, 1994; Grainger & Holcomb, 2009). The model incorporates phonological influences on visual word recognition and the other way around. It assumes that there are bidirectional connections between orthographic and phonological representations on a sublexical and on a lexical level. The shared interface between orthography and phonology is characterized by direct links between orthographically represented words and phonologically represented words and by indirect connections mediated by the sublexical interface (O↔P).

To our knowledge, no study has yet examined how these couplings are best established and whether – given the importance of synchronous retrieval of phonological and orthographic forms during reading (Perfetti, 2007) – the presentational timing of written and spoken forms influences the mapping between orthography and phonology.

We recently addressed this question in two exploratory learning experiments (Funk et al., 2024). Our participants learned Chinese spoken and written (using the Pinyin notation) word forms across three sessions. Critically, the word forms were either presented in synchrony or in asynchronous fashion (spoken preceding or following written form). Inspired by the work by Breznitz and Misra (2003) and

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

by Perfetti's (2007) synchronicity hypothesis, we reasoned that synchronous presentation during learning could benefit the acquisition and mapping of both word forms and would lead to better retention. The opposite was the case. Asynchronous presentation during training generally led to better retention at test compared to synchronous presentation. Indeed, while recall performance was best for spoken-first presentation, it was worst for synchronous presentation. The spoken-last condition lay in between.

Although unexpected in the context of the 'synchronicity hypothesis', several explanations might account for the observed effects. One possibility is that synchronous presentation, which requires simultaneous auditory and visual processing, presents a dual-task situation that was – compared to both asynchronous conditions – more challenging for the participants. Indeed, when processing and acquiring novel information, working memory serves as a bottleneck, according to the Cognitive Load Theory (for review see Mutlu-Bayraktar et al., 2019). Since working memory capacity is limited, perceptual overload adversely affects learning. In multimodal learning settings, processing the same information from bimodal sources reduces the capacity and can thus be more challenging than unimodal settings.

The mapping of spoken and written word forms (i.e., linking at segmental and suprasegmental levels) resulting from exposure in the synchronous condition appeared to be weaker than that resulting from the asynchronous conditions. Another complementary explanation is that in the synchronous condition participants could not engage in cross-modal priming of upcoming structures. That is, as has been proposed in prediction-based theories of language acquisition and processing (e.g., Chang et al., 2006; Dell & Chang, 2014; Reuter et al., 2019), learners try to predict upcoming words to support novel word learning. According to these theories, learners interdependently use correct and incorrect predictions to support language learning. Either the correct predictions are used to reinforce the internal representation or the discrepancy between predicted and encountered information (i.e., prediction error) is used to revise the mental representations and/or the next predictions. A similar mechanism could have supported learning in the asynchronous, but not in the synchronous condition.

To elucidate the mechanisms underlying the behavior observed in our learning experiments, the present work zoomed in on the two explanations above. Specifically, using a cross-modal identity priming paradigm, we sought to quantify how presentational timing, operationalized as varying levels of temporal overlap between spoken and written word forms, affect the spread of activation across phonological and orthographic levels of representation. We did so by quantifying the size of the priming effect induced by each level of temporal overlap.

The Present study

Cross-modal priming paradigms, frequently used to study word recognition (e.g., Hendrickson et al., 2022; Holcomb et al., 2005; Holcomb & Anderson, 1993; Marslen-Wilson & Zwitserlood, 1989), provide evidence for strong interactions between phonological and orthographic representations, showing how information from one modality influences word processing in another (Chng et al., 2019). The direction of the influence depends on a number of factors: According to Slowiaczek and Hamburger (1992), phonological inhibition is interpreted as a manifestation of lexical competition between (phonologically) overlapping words. Facilitation, however, is interpreted as enhanced processing fueled by the overlap at the sublexical level. The special case of repetition priming refers to facilitated processing of words preceded by identical words in the same or in a different modality (i.e., complete segmental/phonological overlap).

Here, we adapted a version where the playback of spoken word primes was shifted relative to the presentation of written targets. It was shifted in three steps: in the synchronous condition, onset of spoken primes and onset of written targets coincided. In the partially overlapping condition, the onset of the written targets was timed to start at nucleus offset of the spoken primes. Finally, in the asynchronous condition, the primes preceded the targets in their entirety such that the onset of the written targets was timed to start at the offset of the recording of the spoken primes[1]. On half of the trials, the written targets were identical to the spoken primes; on the other half, the targets were written pseudowords created by manipulating onset, nucleus, or coda in the existing prime words. Participants carried out a speeded lexical decision task on the written targets. We investigated how different levels of temporal overlap affect participants' lexical decision times to written word and pseudoword targets.

For the synchronous and the asynchronous condition our predictions were relatively straightforward: Recognizing the spoken words prior to written target word onset should result in shorter lexical decision times compared to the synchronous condition where no priming took place. We reasoned that the same general logic should apply to lexical decision times for pseudowords. Hearing the spoken primes in their entirety fully activates their phonological code, activation spreads to associated orthographic representations, which should facilitate the rejection of the mismatching pseudoword targets, compared to the synchronous condition where orthographic representations are not pre-activated.

A special focus of the present experiment was on the 'partially overlapping' condition and how responses to these

[1] Note that we did not include a condition where the written target was not preceded by a spoken prime. In hindsight, such a unimodal baseline would have been useful to evaluate whether reaction times in the synchronous condition were faster (reflecting facilitation) or slower (reflecting inhibition) than reaction times in the target-only condition. However, since our focus was on quantifying the size of the priming effects across multimodal conditions (similar to our learning study), we used the synchronous condition as baseline and asked how the temporal shift of spoken prime presentation affected reaction times to written word and pseudoword targets.

Table 1: Descriptive statistics of controlled experimental lists.

List	*PW manipulation*	*Word frequency*	*PND*	*PNF*	*OND*	*ONF*
1	o: 13; n: 13; c: 14	53.6 (±91.1)	13.2 (± 5.2)	182.5 (± 280.5)	6.4 (± 3.5)	319.3 (± 767.7)
2	o: 13; n: 13; c: 14	41.1 (± 97.9)	13.5 (± 5.3)	213.9 (± 375.3)	6.7 (± 4)	280.2 (± 602.8)
3	o: 13; n: 14; c: 13	67.4 (± 99)	13.2 (± 5.7)	237.1 (± 466.4)	7.4 (± 3.6)	176.2 (± 436.8)
4	o: 13; n: 14; c: 13	43.2 (± 85.4)	13.5 (± 5.5)	213.5 (± 261.4)	6.0 (± 3.3)	222.2 (±356.6)
5	o: 14; n: 13; c: 13	67 (± 155.2)	13.3 (± 6.0)	167.7 (± 190.9)	6.9 (± 3.6)	182.3 (± 239.3)
6	o: 14; n: 13; c: 13	77.7 (± 160.5)	14.4 (± 6.2)	261.3 (± 453.9)	7.6 (± 3.4)	318.0 (± 680.4)

PW – pseudoword; o - onset manipulation for pseudoword generation; n – nucleus manipulation; c – coda manipulation; PND - phonological neighborhood density; PNF – phonological neighborhood frequency; OND – orthographic neighborhood density; ONF – orthographic neighborhood frequency

Values were retrieved from CLEARPOND (Marian et al., 2012) with neighborhood density referring to the sum of addition, deletion, and subtraction neighbors, and the neighborhood frequency referring to the mean frequency thereof.

trials differ from responses to the other two conditions. Presenting part of the spoken primes (i.e., the onset) before the target should prime parts of the orthographic code and therefore facilitate lexical decisions to both words and pseudowords, compared to the synchronous condition (Marslen-Wilson & Zwitserlood, 1989; Slowiaczek & Hamburger, 1992).

Whatmough et al. (1999) showed that simultaneously presenting congruent written word forms slowed down lexical decision to written targets. The authors argued that these effects arose due to differences in the speed with which written and spoken inputs are processed. Considering the structure of BIAM, this means that in addition to the orthographic representation that is activated from the visual target input, orthographic information is activated through the spoken words but with temporal delay. As a result, orthographic candidates are activated simultaneously but at different points in time. Language users then experience competition from target word candidates at different points in time across the two modalities. If spoken word forms would be presented sufficiently in advance of the visual form, visual word recognition should be facilitated as is the case for the asynchronous condition. The open question was whether the interval between prime and target onset in the partially overlapping condition is long enough for activation of the orthographic representations to cascade through the system and facilitate written target recognition (i.e., reaction times).

Concerning the link between the present experiment and our word learning study (Funk et al., 2024), we reasoned that differences in the size of the priming effects relate to differences in learning performance. We return to this issue in the Discussion.

Methods

Participants

Inspired by the sample size recommendations by Brysbaert (2019), we recruited 48 participants (36 self-reported as females; 12 self-reported as males; mean age = 25.13; *SD* = 4.49; range = 18-35) to take part in the present experiment, which featured a fully balanced within-participants design. All participants were native German speakers (43 monolinguals, five bilinguals) with normal hearing, vision, and no known neurological disorder. All participants gave written informed consent to take part in the study and were paid 5€ compensation. Permission to conduct the study was provided by the ethics committee of the German Linguistic Society (DGfS) in accordance with the declaration of Helsinki.

Material

A total of 240 monosyllabic German nouns were selected as word stimuli; eight additional nouns were included as practice trials. For each word, a corresponding pseudoword was generated by manipulating the onset, nucleus or coda of that word such that the resulting pseudoword adhered to the orthographic and phonological rules of German. The words were spoken by a female native speaker of German and were recorded with a sampling frequency of 48 kHz, 16-bit resolution. The spoken words were on average 558 ms long (*SD* = 92; range = 347-798). We used Praat (Boersma & Weenink, 2023) to annotate nucleus offset in each word, which – on average – occurred after 320 ms (*SD* = 100; range = 136 – 574). We then created six experimental lists, which were matched on pseudoword manipulation (onset, nucleus, or coda), word frequency, phonological and orthographic neighborhood density, and neighborhood frequency (Table 1). Specifically, we rotated each word across the three presentational timing conditions and across the two levels of lexicality: On list 1, an existing written target word was preceded by a matching spoken prime. On list 2, the same spoken prime and written target partially overlapped and on list 3, both word forms were presented simultaneously. On list 4, the existing spoken word preceded a pseudoword. On list 5, the existing prime and pseudoword target partially and, on list 6, fully overlapped.

Procedure

Participants were tested individually in a sound-proof booth. Each participant was assigned to one of the experimental

lists, which consisted of eight practice and 240 experimental trials. The 240 experimental trials consisted of 120 matching (existing word prime—same word target) and 120 mismatching (existing word prime—pseudoword target) trials. The order of trials was pseudo-randomized for each participant using *mix* (Van Casteren & Davis, 2006) such that targets with the same lexicality (word, pseudoword) did not occur more than three times in a row and such that the same presentational timing manipulation did not occur more than twice in a row.

Participants were instructed to indicate as quickly as possible whether the presented written target was an existing German word by pressing the respective button on an RTBox (Li et al., 2010). Yes-responses were provided with the right-hand and no-responses with the left-hand thumb.

Each trial started with a central fixation cross presented for 300 ms and was terminated by the participant's button press. The inter-trial interval was 1500 ms. On asynchronous trials, the fixation cross was followed by the playback of the auditory word prime, which was followed by the presentation of the written target (offset prime = onset target). On partially overlapping trials, target onset coincided with the offset of the nucleus in the spoken word primes. On synchronous trials, spoken word prime and written target were presented simultaneously. The experiment lasted approximately 12 minutes.

Analysis

Participants' RTs were calculated as the difference between written target onset and button press. Since the RT distribution was heavily skewed, we log-transformed the data. Only correct responses were included in the analysis. Participants were included if they retained more than 80% of the trials after the exclusion of incorrect responses and responses that were more than ±2 standard deviations away from their log-transformed mean RT. None of the participants was excluded. A total of 10708 data points (93%) was included in the analysis.

The statistical analysis, linear mixed-effects models (LMMs) using the *lme4* package (Bates et al., 2015), was conducted in R (version 4.2.2; R Core Team, 2022). Presentational *Timing* (async, partial overlap, sync), *Lexicality* (word, pseudoword), and their interaction were added as fixed factors to the full LMM. Both factors were dummy-coded with sync and word as reference levels. *Manipulation* was added as further predictor nested within lexicality (only pseudowords were manipulated), to test for effects the manipulation of onset, nucleus, and coda might have had. The model further included Participant and Item as random effects (both with random intercepts). Including random slopes for *Timing* and *Lexicality* by-item and by-participant resulted in the following formula:

*log(rt) ~ timing * lexicality + (1 | lexicality/manipulation) + (1 + timing + lexicality | participant) + (1 + timing + lexicality | item)*

Starting from this model, we used the *buildmer* package (Voeten, 2023), which automatically determines the maximal model that still converges. Using *buildmer*, we then performed backwards stepwise elimination based on likelihood-ratio tests by removing terms that did not reach statistical significance.

Pairwise contrasts and comparisons on the final model were conducted to assess the significance of differences between the levels of *Lexicality* and *Timing* using estimated marginal means obtained using the *emmeans* package (Lenth et al., 2023).

Results

Table 2 and Figure 1 present the descriptive statistics for each of the six conditions. As can be seen, RTs for pseudowords were generally longer than for existing words. Moreover, targets preceded by an asynchronously presented prime were responded to faster than targets that partially overlapped with primes, which were responded to faster than targets presented in synchrony with the corresponding prime.

The results of the linear mixed-effects model analysis are summarized in Figure 2 and Table 3. Taking the maximal feasible model as a starting point, *Manipulation* and the random slope for *Timing* by-participant were dropped through stepwise elimination.

Table 2: Mean RT in ms for each condition. Standard deviation in brackets.

Lexicality/ Timing	Word	Pseudoword
Asynchronous	504 (± 155 *SD*)	583 (± 180 *SD*)
Partially overlapping	562 (± 156 *SD*)	630 (± 181 *SD*)
Synchronous	646 (± 160 *SD*)	734 (± 213 *SD*)

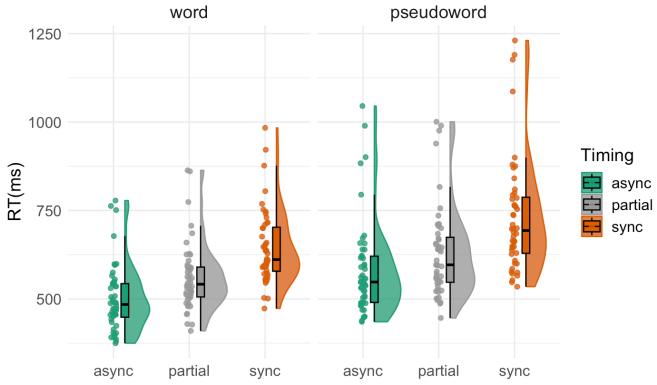

Figure 1: RTs in the LDT. The plot visualizes individual participants' performance (dots) and respective barplots and density functions over all participants.

Table 3: Model summary of log-transformed RT

Predictors	*Estimates*	*SE*	*CI*	*t*	*p*
(Intercept)	6.45	.02	6.40 – 6.49	279.08	<.001
Timing [async]	-.27	.01	-.28 – -.25	-39.66	<.001
Timing [partial]	-.15	.01	-.16 – -.13	-21.31	<.001
Lexicality [pseudoword]	.12	.01	.09 – .15	7.83	<.001
Timing [async] × Lexicality [pseudoword]	.04	.01	.02 – .05	4.13	<.001
Timing [partial] × Lexicality [pseudoword]	-.00	.01	-.02 – .01	-0.33	.745
ICC	0.51				
Marginal R^2	0.177				
Conditional R^2	0.593				
AIC	-5203.373				

The results revealed a main effect for *Lexicality*: Overall, words were responded to faster than pseudowords. We also observed a main effect for *Timing*: Overall, responses to asynchronous trials were faster than responses to partially overlapping and synchronous trials. Post-hoc simple contrasts for *Timing* revealed significant differences for all contrasts within word and pseudoword conditions ($p < .001$).

Finally, we observed an interaction between *Timing* and *Lexicality*, which is visualized in Figure 2. Pairwise comparisons of the priming effects (i.e., the differences between RTs to words and pseudowords of each presentational timing condition revealed significant differences between the asynchronous and the partially overlapping (Z = -4.48; $p < .001$) and between the synchronous and asynchronous (Z = 4.13; $p < .001$) conditions. Crucially, there was no significant difference in the size of the priming effect between the synchronous and the partially overlapping conditions (Z = -.33; p = .74).

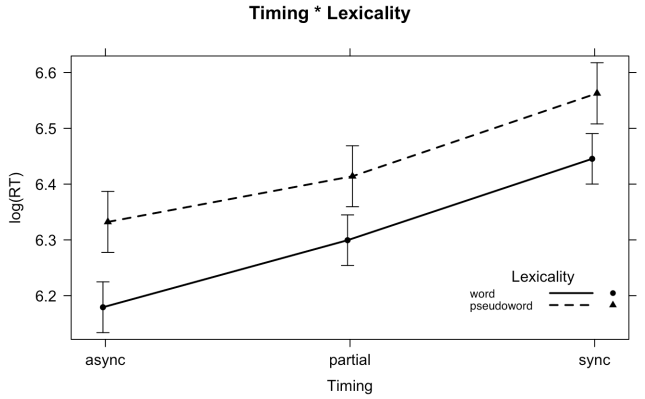

Figure 2: Mixed-effects model results for log(RT) in LDT.

Discussion

The present study used a cross-modal priming paradigm to investigate how different levels of temporal overlap between spoken and written word forms affect lexical decision times to word and pseudoword targets. Our analysis revealed evidence for a main effect of *Lexicality* suggesting that existing words were responded to faster than pseudowords. This effect was present in all three presentational timing conditions. The analysis further revealed a main effect of *Timing*, with fastest responses to word and pseudoword targets preceded by a fully asynchronous prime and slowest responses to targets where prime and target onsets coincided. Response times to targets in the partially overlapping condition lay in between these two conditions.

Our results thus provide clear evidence for cross-modal identity priming in the asynchronous condition. That is, presenting the spoken primes before the written targets led to shorter RTs for both words and pseudowords compared to the partially overlapping and synchronous conditions. This is in line with the assumptions of the bimodal interactive activation model (BIAM) and suggests that activation of the phonological representations spreads to associated orthographic representations, which facilitated lexical decisions (e.g., Diependaele et al., 2010). Similarly, presenting only the prime words' onsets prior to the written targets sped up word and pseudoword responses compared to the synchronous condition, meaning that the activation of the primes' onsets (i.e., phonological representations) partially pre-activated the written targets via sublexical interactive connections (e.g., Grainger et al., 2003; Marslen-Wilson & Zwitserlood, 1989).

As hypothesized, we also observed a difference in RTs for words and pseudowords in the synchronous condition. Since the onset of spoken primes and written targets coincided, this RT difference does not reflect priming. Instead, the RT difference is more likely to reflect a general *Lexicality* effect, based on the observation that rejecting a stimulus as non-lexical takes longer than accepting a stimulus as lexical (e.g., Barca & Pezzulo, 2012). It is important to note that this *Lexicality* effect also contributed to the difference between word and pseudoword response times in the other presentational timing conditions. As described above, the effect size of the word-pseudoword comparisons did not differ between partially overlapping and synchronous conditions, which suggests that the contribution of form priming to the difference between word and pseudoword responses in the partially overlapping condition was minimal (Figure 2).

Although shifting the onset of the spoken primes relative to the written targets, generally facilitated lexical decision times, the level of temporal overlap had differential effects on word and pseudoword responses (i.e., interaction between *Lexicality* and *Timing*). Specifically, our results (see Figure 2) seem to suggest that asynchronous presentation of prime words negatively affected RTs to pseudoword targets, which – compared to partially overlapping and synchronous conditions – resulted in a larger priming effect. One may have predicted the opposite: Being presented with the spoken prime in its entirety fully activated the prime's phonological representation and provided sufficient time for activation to spread to associated orthographic representations. Thus, when encountering a written target that mismatched the primed structure one should be able to reject this stimulus quickly as a pseudoword. A possible account for the results in the asynchronous condition is that the spoken words primed the target phonological and orthographic representations such that on word trials participants were able to accept the written targets quickly, possibly without engaging in 'deep' lexical processing and phonological recoding. This is in line with assumptions of the BIAM that orthographic codes are already activated via interconnections from the spoken input to orthographic representations. On pseudoword trials, on the other hand, participants encountered an orthographic form that mismatched the primed target structure. This mismatch triggered an attempt to process the letter string visually (i.e., a decision based on shallow graphemic processing was not possible), which engaged the process of phonological recoding, which in turn led to increased RTs for pseudoword targets.

In sum, our results suggest that presenting the spoken primes, albeit the words' onset, before the written targets reduces lexical decision times to both words and pseudowords. Asynchronous presentation of the spoken primes resulted in the largest difference between word and pseudoword response times. Overall, the results support mechanisms proposed in the bimodal interactive activation model that encountered spoken words activate an orthographic code which supports cross-modal priming of visual word recognition.

Links between cross-modal priming and word form learning

Do elongated RTs for pseudowords in the asynchronous condition relate to the learning advantage of the asynchronous training condition in our learning study (Funk et al., 2024)? As explained in the Introduction, on asynchronous learning trials, participants first heard novel spoken Chinese word forms and then read the corresponding written forms. That is, they could – based on the spoken input – generate predictions about the upcoming orthographic code and use the discrepancy between predicted and encountered information (i.e., prediction error) to revise their mental representations or in the case of correct predictions reinforce their mapped representations (see Reuter et al., 2019). The size of the priming effect (difference between word and pseudoword response times) in the asynchronous condition in the present experiment suggests that encountering mismatching spoken (i.e., prime words) and visual forms is, as discussed above, associated with enhanced cognitive processing. We speculate that the mechanisms underlying the acquisition (earlier study) and processing of cross-modal (pseudo)word forms (present experiment) do at least partially have a shared origin. That is, just as participants in the present experiment encountered a pseudoword that mismatched the phonologically primed orthographic structure, participants in the earlier study encountered a written Chinese word form that mismatched their auditory-based prediction. Both cases required additional processes, which – in the present study – increased RTs and, in the learning study, supported learning (Reuter et al., 2019). Clearly, more research is needed to follow up on this conjecture. We are currently setting up a follow-up experiment, which utilizes the present cross-modal priming paradigm, and uses matching and mismatching pseudoword pairs. Rather than carrying out a lexical decision task, participants will be instructed to quickly indicate whether spoken prime and visual target are identical. In doing so, we reduce the influence of lexicality on RTs to further disentangle the complex interactions between different levels of representation during multimodal word learning and processing.

References

Barca, L., & Pezzulo, G. (2012). Unfolding visual lexical decision in time. *PLoS ONE*, *7*(4), e35932. https://doi.org/10.1371/journal.pone.0035932

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48.

Boersma, P., & Weenink, D. (2023). *Praat: Doing phonetics by computer [Computer program]* (6.0.33) [Software].

Breznitz, Z., & Misra, M. (2003). Speed of processing of the visual–orthographic and auditory–phonological systems in adult dyslexics: The contribution of "asynchrony" to word recognition deficits. *Brain and Language*, *85*(3), 486–502. https://doi.org/10.1016/S0093-934X(03)00071-3

Brysbaert, M. (2019). How many participants do we have to include in properly powered experiments? A tutorial of power analysis with references tables. *Journal of Cognition*, *2*(1), 1–38. https://doi.org/10.5334/joc.72

Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological Review*, *113*(2), 234–272. https://doi.org/10.1037/0033-295X.113.2.234

Chng, K. Y. T., Yap, M. J., & Goh, W. D. (2019). Cross-modal masked repetition and semantic priming in auditory lexical decision. *Psychonomic Bulletin & Review*, *26*(2), 599–608. https://doi.org/10.3758/s13423-018-1540-8

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A Dual Route Cascaded Model of visual word recognition and reading aloud. *Psychological Review*, *108*(1), 204–256.

Dell, G. S., & Chang, F. (2014). The P-chain: Relating sentence production and its disorders to comprehension and acquisition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1634), 20120394. https://doi.org/10.1098/rstb.2012.0394

Diependaele, K., Ziegler, J. C., & Grainger, J. (2010). Fast phonology and the Bimodal Interactive Activation Model. *European Journal of Cognitive Psychology*, *22*(5), 764–778. https://doi.org/10.1080/09541440902834782

Frost, R. (1998). Toward a strong phonological theory of visual word recognition: True issues and false trails. *Psychological Bulletin*, *123*(1), 71–99.

Funk, J., Huettig, F., & Hintz, F. (2024). The role of presentational timing in acquiring novel written and spoken word forms. *Journal of Experimental Psychology: General*. https://doi.org/10.17605/OSF.IO/7AT9D

Grainger, J., Diependaele, K., Spinelli, E., Ferrand, L., & Farioli, F. (2003). Masked repetition and phonological priming within and across modalities. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(6), 1256–1269. https://doi.org/10.1037/0278-7393.29.6.1256

Grainger, J., & Ferrand, L. (1994). Phonology and orthography in visual word recognition: Effects of masked homophone primes. *Journal of Memory and Language*, *33*, 218–233.

Grainger, J., & Holcomb, P. J. (2009). Watching the word go by: On the time-course of component processes in visual word recognition. *Language and Linguistics Compass*, *3*(1), 128–156. https://doi.org/10.1111/j.1749-818X.2008.00121.x

Harm, M. W., & Seidenberg, M. S. (2004). Computing the meanings of words in reading: Cooperative division of labor between visual and phonological processes. *Psychological Review*, *111*(3), 662–720. https://doi.org/10.1037/0033-295X.111.3.662

Hendrickson, K., Apfelbaum, K., Goodwin, C., Blomquist, C., Klein, K., & McMurray, B. (2022). The profile of real-time competition in spoken and written word recognition: More similar than different. *Quarterly Journal of Experimental Psychology*, *75*(9), 1653–1673. https://doi.org/10.1177/17470218211056842

Holcomb, P. J., & Anderson, J. E. (1993). Cross-modal semantic priming: A time-course analysis using event-related brain potentials. *Language and Cognitive Processes*, *8*(4), 379–411. https://doi.org/10.1080/01690969308407583

Holcomb, P. J., Anderson, J., & Grainger, J. (2005). An electrophysiological study of cross-modal repetition priming. *Psychophysiology*, *42*(5), 493–507. https://doi.org/10.1111/j.1469-8986.2005.00348.x

Lenth, R. V., Bolker, B., Buerkner, P., Giné-Vázquez, I., Herve, M., Jung, M., Love, J., Miguez, F., Riebl, H., & Singmann, H. (2023). *emmeans: Estimated marginal means, aka least-squares means* [Software].

Li, X., Liang, Z., Kleiner, M., & Lu, Z.-L. (2010). RTbox: A device for highly accurate response time measurements. *Behavior Research Methods*, *42*(1), 212–225. https://doi.org/10.3758/BRM.42.1.212

Marian, V., Bartolotti, J., Chabal, S., & Shook, A. (2012). CLEARPOND: Cross-linguistic easy-access resource for phonological and orthographic neighborhood densities. *PLoS ONE*, *7*(8), e43230. https://doi.org/10.1371/journal.pone.0043230

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, *25*(1–2), 71–102. https://doi.org/10.1016/0010-0277(87)90005-9

Marslen-Wilson, W. D., & Zwitserlood, P. (1989). Accessing spoken words: The importance of word onsets. *Journal of Experimental Psychology: Human Perception and Performance*, *15*(3), 576–585.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*(1), 1–86. https://doi.org/10.1016/0010-0285(86)90015-0

McCusker, L. X., Hillinger, M. L., & Bias, R. G. (1981). Phonological recoding and reading. *Psychological Bulletin*, *89*(2), 217–245.

Mutlu-Bayraktar, D., Cosgun, V., & Altan, T. (2019). Cognitive load in multimedia learning environments: A systematic review. *Computers & Education*, *141*, 103618. https://doi.org/10.1016/j.compedu.2019.103618

Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading*, *11*(4), 357–383.

Perre, L., Midgley, K., & Ziegler, J. C. (2009). When *beef* primes *reef* more than *leaf*: Orthographic information affects phonological priming in spoken word recognition. *Psychophysiology*, *46*(4), 739–746. https://doi.org/10.1111/j.1469-8986.2009.00813.x

Rastle, K., & Brysbaert, M. (2006). Masked phonological priming effects in English: Are they real? Do they matter? *Cognitive Psychology*, *53*(2), 97–145. https://doi.org/10.1016/j.cogpsych.2006.01.002

Rastle, K., McCormick, S. F., Bayliss, L., & Davis, C. J. (2011). Orthography influences the perception and production of speech. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(6), 1588–1594. https://doi.org/10.1037/a0024833

Reuter, T., Borovsky, A., & Lew-Williams, C. (2019). Predict and redirect: Prediction errors support children's word learning. *Developmental Psychology*, *55*(8), 1656–1665. https://doi.org/10.1037/dev0000754

Slowiaczek, L. M., & Hamburger, M. B. (1992). Prelexical facilitation and lexical interference in auditory word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(6), 1239–1250.

Taylor, J. S. H., Davis, M. H., & Rastle, K. (2017). Comparing and validating methods of reading instruction using behavioural and neural findings in an artificial orthography. *Journal of Experimental Psychology: General*, *146*(6), 826–858. https://doi.org/10.1037/xge0000301

Van Casteren, M., & Davis, M. H. (2006). Mix, a program for pseudorandomization. *Behavior Research Methods*,

38(4), 584–589. https://doi.org/10.3758/BF03193889

Voeten, C. C. (2023). *buildmer: Stepwise elimination and term reordering for mixed-effects regression* [Software]. https://cran.r-project.org/package=buildmer

Whatmough, C., Arguin, M., & Bub, D. (1999). Cross-modal priming evidence for phonology-to-orthography activation in visual word recognition. *Brain and Language*, *66*(2), 275–293. https://doi.org/10.1006/brln.1998.1996

Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin*, *131*(1), 3–29. https://doi.org/10.1037/0033-2909.131.1.3