*Applied Psycholinguistics* (2026), **47**, e37, 1–23
doi:10.1017/S0142716426100812

CAMBRIDGE UNIVERSITY PRESS

ORIGINAL ARTICLE

# Modality-specific orthographic priming in Chinese semantic processing

Tiande Huang[1] and Hengshuang Liu[2]

[1]Faculty of European Languages and Cultures, Guangdong University of Foreign Studies, China and [2]Bilingual Cognition and Development Lab, Centre for Linguistics and Applied Linguistics, Guangdong University of Foreign Studies, China
**Corresponding author:** Hengshuang Liu; Email: liuhengshuang@gdufs.edu.cn

(Received 9 November 2025 UTC; revised 28 May 2026 UTC; accepted 3 June 2026 UTC)

**Abstract**
This study examines how orthographic knowledge modulates semantic processing across auditory and visual modalities in Chinese, a nonalphabetic script with opaque sound-spelling mappings. Forty-eight native Mandarin speakers performed semantic-relatedness judgements on disyllabic word pairs. Orthographic overlap (shared vs. non-shared character) and common syllable position (first vs. second) were systematically manipulated across auditory and visual sessions, with word frequency, word class, and other confounds carefully controlled. Results revealed a striking cross-modal dissociation. Robust orthographic facilitation emerged in auditory processing but was absent in visual processing. In audition, this facilitation was substantially stronger when the shared character occurred in the second syllable, where phonological disambiguation demands peak, than in the first syllable. Additional modality-specific patterns emerged: target word frequency selectively modulated reaction times in audition, whereas cross-class orthographic overlap unexpectedly produced response interference in vision. These findings indicate that orthographic activation is selectively recruited during Chinese spoken word recognition to resolve lexical ambiguity during temporally extended auditory input. The results support interactive activation models while underscoring that lexical access pathways are both modality-specific and script-specific.

**Keywords:** Chinese spoken word recognition; cross-modal processing; interactive activation model; orthographic priming; syllable position effects

## Introduction

The semantic processing of spoken language, while being the most frequently utilized modality in daily communication, has received considerably less research attention compared to reading comprehension (Rodd et al., 2005). A growing body of evidence suggests that learning to read and write can fundamentally alter the cognitive architecture involved in auditory language processing (Frith, 1998).

© The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/S0142716426100812 Published online by Cambridge University Press

This claim is strongly supported by studies demonstrating that spoken word recognition is influenced by orthographic knowledge, manifesting in effects of orthographic neighborhood density (Ziegler et al., 2003), orthographic consistency (Ventura et al., 2004; Ziegler & Ferrand, 1998), and orthographic similarity between words (Chéreau et al., 2007; Pattamadilok et al., 2008; Seidenberg & Tanenhaus, 1979; Slowiaczek et al., 2003; Taft et al., 2008; Ziegler et al., 2008). These orthographic effects have been reliably observed across a diverse range of experimental paradigms, including rhyme judgment (e.g., Seidenberg & Tanenhaus, 1979), phoneme monitoring (e.g., Frauenfelder et al., 1990), primed auditory lexical decision (e.g., Chéreau et al., 2007), semantic categorization (e.g., Pattamadilok et al., 2009), and even nonlinguistic noise detection tasks (e.g., Perre et al., 2011).

The pervasive finding of orthographic mediation in auditory word processing, irrespective of the specific orthographic property manipulated or the auditory task employed, provides compelling support for interactive activation or connectionist models of lexical access (Grainger & Ferrand, 1996; Muneaux & Ziegler, 2004; Plaut et al., 1996). These models posit the existence of bidirectional, symmetric connections between phonological and orthographic representations, forming an amodal lexical network where activation can flow interactively across modalities.

It is crucial to note, however, that the vast majority of this evidence originates from studies conducted in alphabetic languages such as English (Chéreau et al., 2007; Miller & Swick, 2003; Slowiaczek et al., 2003; Taft et al., 2008), Portuguese (Ventura et al., 2004), French (Pattamadilok et al., 2007; Ziegler & Ferrand, 1998; Ziegler et al., 2008), and Thai (Pattamadilok et al., 2008). In these writing systems, the mapping between spelling/graphemes and sound/phonemes is relatively systematic and transparent. Consequently, observing an influence of orthographic knowledge on phonological processing is, to some extent, less surprising. The critical and unresolved question is whether such orthographic effects persist in a nonalphabetic language like Chinese, where the relationship between a character's form and its pronunciation is largely arbitrary and opaque.

One might hypothesize that the profound disparity between orthography and phonology in Chinese would minimize the potential for orthographic variables to influence spoken word recognition. Conversely, a strong case can be made that orthographic mediation might be even more critical for disambiguation in Chinese auditory comprehension. This argument rests on three unique characteristics of the Chinese language.

First, meaning is inherently less ambiguous in written Chinese than in its spoken form due to the exceptionally high rate of homophony at the character level. Among high-frequency characters, 5,265 are represented by only 1,277 distinct syllables when tonal differences are considered (Chao, 1976; Yin & Rohsenow, 1994). This means a single spoken syllable rarely maps uniquely to one character; approximately 55% of syllables correspond to five or more distinct homophonous characters (Treiman et al., 1981). For instance, the syllable /ye4/[1] corresponds to multiple characters with unrelated meanings, such as 叶(*leaf*), 页(*page*), 业(*career*), and 夜(*night*). Therefore, in situations where contextual cues are minimal or absent, such as in isolated word recognition paradigms common in psycholinguistic research, the mental activation of a word's orthographic form may be necessary to resolve this pervasive homophony.

https://doi.org/10.1017/S0142716426100812 Published online by Cambridge University Press

Second, the visual form of Chinese characters often provides direct semantic cues. Many characters originated as pictographs representing concrete objects (e.g., 日/ri4/ *sun* and 月/yue4/ *moon*) or ideographs formed by association or analogy (e.g., duplicating 人/ren2/ *person* to form 从/cong2/ *to follow*) (Feng et al., 2001). Even in modern Chinese, semantic radicals, which hint at a character's meaning, are embedded in approximately 85% of characters (Zhu, 1988). For example, the character 树 (/shu4/ *tree*) contains the semantic radical 木 (/mu4/ *wood*). This deep interconnection between orthography and semantics makes it plausible for literate individuals to infer meaning directly from the written form, a tendency that may carry over into auditory processing.

Third, while phonetic components in characters can provide some phonological hints, they reliably cue the correct pronunciation in only about 38% of the characters in which they appear (Zhou, 1978). Consequently, Chinese writing does not reflect the segmental phonological structure of the language, and the sound-spelling mapping is highly arbitrary. This fundamental opacity offers a unique advantage: it allows for a clearer dissociation of orthographic effects from phonological effects than is typically possible in alphabetic languages (Paap & Noel, 1991), where the two are often confounded.

These cross-linguistic differences underscore the theoretical importance of investigating orthographic involvement in spoken word recognition within Chinese. To date, five empirical studies have directly addressed this issue, yielding mixed results (Chen et al., 2016; Qu & Damian, 2016; Zhou & Marslen-Wilson, 1995; Zou et al., 2012; Zou et al., 2015). Most of these studies compared reaction times to targets preceded by primes that shared a whole character (e.g., 剧烈/ju4 lie4/ *violent* & 剧本/ju4 ben3/ *play script*) versus primes that were only phonologically identical (e.g., 惧怕/ju4 pa4/ *fear* & 剧本/ju4 ben3/ *play script*) (Zhou & Marslen-Wilson, 1995; Zou et al., 2012; Zou et al., 2015). However, two recent studies employed different approaches. Qu and Damian (2016) manipulated orthographic similarity at the sub-character level by ensuring that word pairs shared identical phonetic radicals in one character, rather than sharing a whole character. In contrast, Chen et al. (2016) did not use a priming paradigm at all; instead, they asked participants to judge whether a spoken word, whose orthographic consistency and homophone density were manipulated, referred to an animal.

The inconsistent findings across these studies likely stem from several methodological variations. First, the use of a lexical decision task (e.g., Zhou & Marslen-Wilson, 1995; Zou et al., 2012) may be problematic. Participants could potentially make lexicality judgments based on the familiarity of the auditory word form alone, without engaging in deeper orthographic or semantic processing, thereby masking orthographic effects.

Second, the composition of prime-target pairs differed. Zhou and Marslen-Wilson (1995), who found no orthographic effect, included both semantically related and unrelated pairs. Zou et al. (2012), who reported a significant effect, used only unrelated pairs. This raises the possibility that the observed orthographic effect in the latter study was inflated, as participants might rely more heavily on orthographic information when semantic cues are absent.

Third, a direct comparison between studies by the same research team reveals a puzzling discrepancy. Zou et al. (2012) found an orthographic effect using a lexical

https://doi.org/10.1017/S0142716426100812 Published online by Cambridge University Press

decision task, whereas Zou et al. (2015) did not find one using a morphological relatedness task. It is difficult to determine whether this difference is due to the superficial nature of lexical decision, inadequate control of confounds such as word frequency and word class, or others.

Fourth, Chen et al. (2016) observed orthographic effects in ERP data but not in behavioral measures. The limited influence of orthography on behavior may be explained, on the one hand, by the absence of a priming paradigm in their study, and on the other, by the fact that their proxy for orthographic effects—homophone density—captures not only orthographic competition but also semantic competition strength (Wang et al., 2012).

Fifth, while both Qu and Damian (2016) and Zou et al. (2015) employed semantic judgment tasks, their findings diverged likely due to several factors. (a) Qu and Damian (2016) required a judgment on the semantic relatedness of two disyllabic words, whereas Zou et al. (2015) assessed whether a specific character represented the same morpheme. The latter task may more readily induce strategic visualization of characters, potentially leading to artifactual orthographic effects. (b) In Zou et al. (2015), the word pairs in both compared conditions contained phonologically identical characters. In contrast, Qu and Damian (2016) used word pairs that were phonologically unrelated in both conditions, which introduced a greater risk of contamination from uncontrolled phonetic similarities. (c) Qu and Damian (2016) operationalized orthographic similarity at the sub-character radical level, while Zou et al. (2015) manipulated it at the whole-character level. When only radicals are shared, unintended orthographic similarities, such as similar semantic radicals and overall structural resemblance, can confound the results. In contrast, whole-character identity is a clear, categorical variable that minimizes such confounding influences. Furthermore, in Chinese, where meaning is primarily conveyed by disyllabic compounds rather than single characters (Wang et al., 1986), overlap of a whole character between two words is less likely to be confounded with lexical repetition effects than in languages where words are typically monomorphemic.

Critically, although Zou et al. (2012) systematically manipulated the phonological and orthographic match at the first syllable of disyllabic prime-target pairs, none of the existing studies has examined whether orthographic effects differ when the homophonic syllable occurs in the second syllable. This omission is theoretically significant because models of spoken word recognition posit that auditory input is processed incrementally: upon hearing the first syllable, listeners activate a broad cohort of lexical candidates that share that phonological onset (Marslen-Wilson, 1987; Marslen-Wilson & Welsh, 1978; Norris, 1994). In Chinese, this initial cohort is exceptionally large due to high syllable-level homophony (Treiman et al., 1981; Yin & Rohsenow, 1994). When the second syllable arrives, it provides critical diagnostic information that sharply narrows the candidate set, creating a temporally precise "disambiguation window." At this juncture, the co-activation of a shared orthographic form may supply additional top-down constraint, thereby yielding stronger facilitation than when the overlap occurs in the first syllable, where the cohort remains broad and ambiguous (Dumay et al., 2001; Slowiaczek & Hamburger, 1992). Thus, syllable position is not merely a methodological variable but a theoretically motivated factor rooted in the incremental architecture of auditory lexical access.

https://doi.org/10.1017/S0142716426100812 Published online by Cambridge University Press

Conversely, visual word recognition presents a fundamentally different computational landscape. Because both characters of a disyllabic compound are available simultaneously, readers can map orthographic form directly onto semantic content without relying on incremental disambiguation (Cheng & Shih, 1988; Pollatsek et al., 2000). If orthographic priming is driven by the need to resolve temporal ambiguity in sequential input, then no comparable positional asymmetry should emerge in the visual modality, where orthographic information is holistically present from the outset (Whitney, 1998). This cross-modal dissociation allows for a direct test of whether orthographic facilitation reflects amodal lexical representations, as predicted by interactive activation models (Grainger & Ferrand, 1996; Plaut et al., 1996), or rather arises from modality-specific processing demands tied to the temporal structure of the input.

To clarify the extant literature, the current study implements a semantic relatedness judgment task. This task requires participants to access the meanings of both prime and target words fully, making it unlikely to trigger strategic visual imagery while ensuring deep semantic processing. Crucially, the orthographic overlap is defined as the identity of a whole character between the prime and target, ensuring a clear and well-controlled manipulation that minimizes confounds present in radical-based approaches. All test pairs are phonologically identical in one syllable, and for half of them, this syllable also corresponds to a shared character. The position of the shared syllable (first vs. second) was systematically manipulated to examine whether orthographic facilitation is greater when the shared character occurs in the second syllable, where disambiguation demands are highest, than when it occurs in the first syllable. Furthermore, possible confounding variables, including word frequency, word class, audio length, and cohort density,[2] are carefully matched during stimulus selection and included as covariates in the statistical analysis. Finally, to directly test whether orthographic facilitation is specific to auditory processing or also occurs in visual word recognition, participants perform the task in both auditory and visual modalities. Under such a rigorous experimental design, the present study aims to provide unequivocal evidence regarding the role of orthography in Chinese auditory semantic processing and its contribution to broader theories of interactive activation model.

## Methods

### *Participants*

Based on the current experimental design, G*Power 3.1.9.2 calculations indicated that a minimum of 16 participants would be required to achieve a medium effect size of $f = 0.25$ with a significance level of 0.05 and a statistical power of 0.8. To aim for a stronger effect size and higher statistical power, we recruited 48 participants (26.4 ± 3.0 years, 19–33 years; 24 females). All participants were native Mandarin speakers and use simplified Chinese characters in daily reading and writing. They were self-reported to be highly proficient in Chinese reading (6.69 ± 0.55, 5–7) and listening (6.67 ± 0.60, 5–7) on a 7-point Likert scale (1 = extremely unskilled, 2 = quite unskilled, 3 = slightly unskilled, 4 = neither skilled nor unskilled, 5 = slightly skilled, 6 = quite skilled, 7 = extremely skilled). All participants were

https://doi.org/10.1017/S0142716426100812 Published online by Cambridge University Press

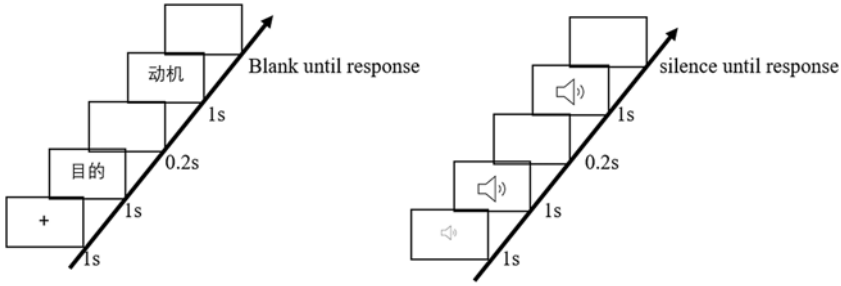



**Figure 1.** Timing of a trial in a visual session (left) or an auditory session (right). The example is a synonymous filler (Prime: 目的 /mu4 di/ *goal*; Target: 动机 /dong4 ji1/ *motivation*).

right-handed, as indicated by the Edinburgh Handedness Inventory (86.0 ± 16.4, 42.9–100) (Oldfield, 1971). None of the participants had a history of psychiatric or neurological disorders. Information such as age, gender, handedness, psychological health, and language background was collected using a demographic questionnaire before the experiment. Ethical approval was obtained from the relevant Institutional Review Board (IRB), and informed consent was obtained prior to the start of the experiment.

### *Procedure*

A semantic relatedness judgment task was administered in separate visual and auditory sessions. The sequence of the two sessions was counterbalanced among participants. During each session, participants had to complete a 6-trial practice prior to the 40-trial formal experiment. Each trial began with either a blue cross in the screen center (in the visual session) or a 700 Hz beep (in the auditory session), presented for 1 s to signal the start of a trial, followed by a pair of two-character words displayed consecutively for 1 s each with a 0.2 s inter-stimulus interval in between (Figure 1). In the visual paradigm, written words were presented in 80-point Nsimsun font created in bitmap format; in the auditory paradigm, spoken words were a female voice synthesized by Interphonic 5.0 at a 16,000 Hz audio rate and 16-bit audio size in wav format. In each trial, participants were asked to decide whether the sequentially presented words were synonymous via a button press (right index or middle finger for "yes" or "no", respectively). The first word in each pair acted as the prime for the second word, which was the target. Reaction time was recorded from the target onset; the subsequent trial did not begin until a response was recorded.

Each set of four prime-target pairs was constructed around a central word that acted as the prime or the target in each pair. In each set, the test stimuli were two non-synonyms of the central word, both of which shared phonology with the central word, but only one shared orthography. Comparisons between these two test non-synonyms allowed us to investigate the effect of orthography on spoken word processing. To create filler stimuli, two synonyms of the central word were selected from synonym dictionaries (Zhang, 2010; Zhao & Li, 2009), such that one synonym shared orthography and phonology with the central word, and the other shared neither (Table 1). Although the two synonyms are asymmetric compared to the two

https://doi.org/10.1017/S0142716426100812 Published online by Cambridge University Press

**Table 1.** Examples of the four conditions in the semantic relatedness judgment task

| Trial type | Meaning | Orthography | Phonology | Prime | Target |
|---|---|---|---|---|---|
| Test trials | Different | Shared | Shared | 目的 (/mu4 di/ *goal*) | 目录 (/mu4 lu4/ *catalog*) |
| | | Non-shared | Shared | 木头 (/mu4 tou/ *wood*) | 目的 (/mu4 di/ *goal*) |
| Filler trials | Similar | Shared | Shared | 目的 (/mu4 di/ *goal*) | 目标 (/mu4 biao1/ *target*) |
| | | Non-shared | Non-shared | 动机 (/dong4 ji1/ *motivation*) | 目的 (/mu4 di/ *goal*) |

*Note:* Pronunciation and meaning follow each word in parentheses; notably, 的 and 头 are pronounced in a neutral tone (no tone) in 目的 (/mu4 di/ *goal*) and 木头 (/mu4 tou/ *wood*), respectively.

non-synonym ones, they primarily serve as fillers to balance the numbers of synonym and non-synonym pairs and are not used to examine the orthographic effect.

Given the combination of semantics (similar or not), phonology (shared or not), and orthography (shared or not), in theory, it would be possible to create eight different prime-target types. However, the task was limited to four trial types, as the other four types either do not exist (such as word pairs similar in form but not in pronunciation, regardless of whether they are synonyms), are extremely rare (like synonym pairs containing homophones), or, although they do exist, are of little relevance to this study (for instance, word pairs dissimilar in sound, form, and meaning are abundant but neither allow contrast with other categories to probe the orthographic effect nor serve as filler conditions to balance the number of positive and negative responses).

We elected to use two-character rather than single-character words due to their high prevalence in Chinese and low homophony. With regard to prevalence, in a corpus of 1.31 million modern Chinese words, there are more than six times as many two-character words as single-character words (approximately 73.6% and 12%, respectively, Wang et al., 1986). With regard to homophony, two-character compounds are more distinguishable than single-character words in auditory presentation due to a lower incidence of homophones (Zhou & Marslen-Wilson, 1995). For example, while a single spoken syllable /mu4/ could represent multiple single-character words such as 木 (*wood*), 目 (*eye*), and 暮 (*dusk*), the disyllable /mu4 di/ corresponds to only one two-character word, 目的 (*goal*), without ambiguity.

To establish the degree of synonymy for each word pair in the stimulus list, 30 Chinese-speaking participants (15 females; 23.5 ± 3.5 years, 19–30 years; handedness: 86.6 ± 13.1, 63.6–100; reading: 6.67 ± 0.55, 5–7; listening: 6.83 ± 0.46, 5–7) rated each pair of stimuli using a 7-point Likert scale (1 = surely non-synonymous, 2 = should be non-synonymous, 3 = probably non-synonymous, 4 = hard to decide, 5 = probably synonymous, 6 = should be synonymous, 7 = surely synonymous). Results showed that participants perceived paired words in the two synonymous conditions as having significantly higher synonymy than those in the other two non-synonymous conditions (Synonyms: 5.6 ± 0.7; Non-synonyms: 1.1 ± 0.2; Comparison: $t(96.113) = 51.730$, $p < .001$). Importantly, for the two test conditions designed to be non-synonyms, ratings were low, with no significant difference between them (Shared orthography: 1.19 ± 0.32; Non-shared orthography: 1.03 ± 0.043; Comparison: $t(156) = 1.339$, $p = .183$).

https://doi.org/10.1017/S0142716426100812 Published online by Cambridge University Press

In addition to synonymy, other word properties such as the stroke number, the word frequency, and the spatial configuration were all checked to be comparable between responses and within test conditions (Table 2). Word frequency was indicated by word count per million as specified in the SUBTLEX-CH corpus. This corpus is based on 46.8 million Chinese characters and 33.5 million Chinese words coded from film subtitles and television program subtitles and has been validated as providing sound frequency estimates of daily written and spoken Chinese exposure (Cai & Brysbaert, 2010).

Experiments were programmed using E-prime 2.0 (Schneider et al., 2002). The 40 trials in each session were evenly distributed and counterbalanced across the four conditions (10 trials per condition); across the two response types (20 trials each for synonyms and non-synonyms); across the position of the central word in the prime-target pair (20 trials each for prime and target); and across the position of the homophonic syllable in the two-syllable words[3] (20 word pairs overlapping in the first syllable, such as 目的 /mu4 di/ *goal* & 木头 /mu4 tou/ *wood*; 20 word pairs sharing the second syllable, such as 按照/an4 zhao4/ *according to* & 拍照/pai1 zhao4/ *take a photo*). To avoid familiarity effects, the four stimulus pairs sharing a central word were pseudo-randomly assigned between four separate stimulus lists, and only one of the four lists was administered to each participant in a Latin square sequencing. No word pair was displayed more than once in any given list. In the visual or auditory session from a given list, the 40 trials were presented in the same randomized order for all participants, with the first three trials uniformly set as filler conditions to familiarize participants with the task. Each stimulus list took around ten minutes to complete, including both visual and auditory sessions.

### *Data handling*

All analyses were conducted using SPSS 25 in the current study, where reaction time was used to indicate processing complexity and effort (Baayen & Milin, 2010; Lachaud & Renaud, 2011; Whelan, 2008). Reaction times in the visual and auditory sessions were handled separately due to differences in the timing of responses driven by the different presentation modalities: In the visual modality, the whole word form is available simultaneously, allowing for faster reactions to the target, whereas in the auditory modality, one must wait until sufficient information is heard before a decision about synonymy can be made (Harley, 2013; Whitney, 1998). Incorrect responses were excluded from analysis of reaction times. Finally, of the 960 reaction time values per session (10 trials per condition × 2 test conditions × 48 participants), 922 and 919 values were included for analysis of the auditory and visual sessions, respectively.

Reaction times characteristically had a skewed distribution with a long tail. To satisfy the normality assumption of parametric tests, raw reaction times were normalized prior to analysis. To find the transformation that best approximated normality, inverse, logarithm, square root, and cube root transformations were performed. Among them, the inverse transformation was chosen because it outperformed other transformations with a distribution shape most approximating normality. Following inverse transformation, individual extremes were identified using a boxplot, where inverse reaction times at least three times the interquartile range above the 3rd quartile or below the 1st quartile were excluded and retested

https://doi.org/10.1017/S0142716426100812 Published online by Cambridge University Press

https://doi.org/10.1017/S0142716426100812 Published online by Cambridge University Press

**Table 2.** Summary of stimulus properties across conditions

| | Central word | Synonyms with shared phonology and orthography | Synonyms without shared phonology or orthography | Non-synonyms with shared orthography | Non-synonyms without shared orthography | Synonyms versus non-synonyms | Differences between the two test conditions |
|---|---|---|---|---|---|---|---|
| **Synonymy (1–7)** | – | 5.5 | 5.7 | 1.2 | 1.0 | $t(96.1) = 51.730$, $p < .001$ | $t(156) = 0.407$, $p = .183$ |
| **Stroke number** | 16.5 | 16.4 | 16.6 | 17.3 | 16.1 | $t(195) = -0.315$, $p = .753$ | $t(195) = 1.260$, $p = .209$ |
| **Word frequency (per million)** | 5910 | 4510 | 9540 | 5330 | 3320 | $t(67.6) = 1.426$, $p = .158$ | $t(63.4) = 1.319$, $p = .192$ |
| **Spatial configuration** | | | | | | | |
| **1st character, Left-right** | 19 | 20 | 22 | 20 | 16 | $X^2(1) = 0.462$, $p = .497$ | $X^2(1) = 0.444$, $p = .505$ |
| **2nd character, Left-right** | 15 | 19 | 16 | 16 | 19 | $X^2(1) = 0$, $p = 1$ | $X^2(1) = 0.257$, $p = .612$ |
| **1st character, Top-down** | 10 | 14 | 10 | 8 | 14 | $X^2(1) = 0.087$, $p = .768$ | $X^2(1) = 1.636$, $p = .201$ |
| **2nd character, Top-down** | 12 | 12 | 13 | 14 | 10 | $X^2(1) = 0.020$, $p = .886$ | $X^2(1) = 0.667$, $p = .414$ |
| **1st character, Single** | 6 | 4 | 5 | 6 | 7 | $X^2(1) = 0.727$, $p = .394$ | $X^2(1) = 0.077$, $p = .782$ |
| **2nd character, Single** | 7 | 4 | 6 | 2 | 8 | $X^2(1) = 0$, $p = 1$ | $X^2(1) = 3.6$, $p = .058$ |
| **1st character, Half-surrounding** | 5 | 2 | 3 | 5 | 3 | $X^2(1) = 0.692$, $p = .405$ | $X^2(1) = 0.5$, $p = .48$ |
| **2nd character, Half-surrounding** | 6 | 4 | 5 | 6 | 2 | $X^2(1) = 0.059$, $p = .808$ | $X^2(1) = 2.0$, $p = .157$ |

*Note:* The stroke number and the spatial configuration were referenced from an online Chinese dictionary (https://zdic.net/); the word frequency was referenced from the SUBTLEX-CH corpus (Cai & Brysbaert, 2010).

until no further extremes were detected. Finally, one extreme was removed from the visual and the auditory sessions each. After individual exclusion of extremes, the inverse-transformed reaction times were shown to be normally distributed for each combination of conditions by common syllable positions without extremes detected. However, the distribution was no longer normal after back-transforming the inverse reaction times, suggesting that subsequent parametric tests should be conducted on the inverse data rather than the back-transformed data.

### *Statistical analysis*

To provide basic information about task difficulty, a generalized linear mixed model was conducted to test whether participants' accuracy rates were significantly affected by the condition, the common syllable position, the modality, or their interactions, with participant ID and item ID as random intercepts.

Regarding the data processing of reaction times, initially, a unified linear mixed model was fitted to inverse-transformed reaction times with random intercepts for participants and items. Fixed effects included modality, condition, common syllable position (Dumay et al., 2001; Slowiaczek & Hamburger, 1992; Zhou & Marslen-Wilson, 1995), within-pair similarity in word class (based on the dominant word class in SUBTLEX-CH), $\log_{10}$ target word frequency (SUBTLEX-CH; Feng et al., 2001; Harley, 2013; Marslen-Wilson, 1987; Whitney, 1998), and all two- and three-way interactions involving modality and condition. Although most lexical variables were matched during stimulus selection, this initial model was intended primarily to examine the interactions involving modality rather than examining main effects.

Results from this omnibus model revealed a significant three-way interaction of modality × condition × common syllable position, $F(3, 877) = 5.207$, and $p = .001$, supporting the need for modality-specific analyses. Accordingly, we proceeded to fit separate linear mixed models for the visual and auditory sessions to clearly interpret the orthographic priming effect within each modality.

For the auditory session, a linear mixed model was applied to inverse reaction times with random intercepts for participants and items. Fixed effects included condition, common syllable position, within-pair similarity in word class, $\log_{10}$ target word frequency, audio duration, and cohort density of each target word and all two-way interactions with condition.

For the visual session, a separate linear mixed model was conducted on inverse reaction times, including condition, common syllable position, within-pair similarity in word class, $\log_{10}$ target word frequency, and all two-way interactions with condition, again with random intercepts for participants and items.

## Results

Regarding the accuracy, generalized linear mixed model showed that neither the main effect of the modality nor any interactions with the modality was significant. As illustrated in Figure 2, the main effect of the condition ($t = -2.389$, $p = .017$) and the interaction of the condition by the common syllable position ($t = 2.113$, $p = .035$) were both significant. Simple effects analysis revealed that condition did not significantly affect accuracy, irrespective of whether the shared syllable

https://doi.org/10.1017/S0142716426100812 Published online by Cambridge University Press

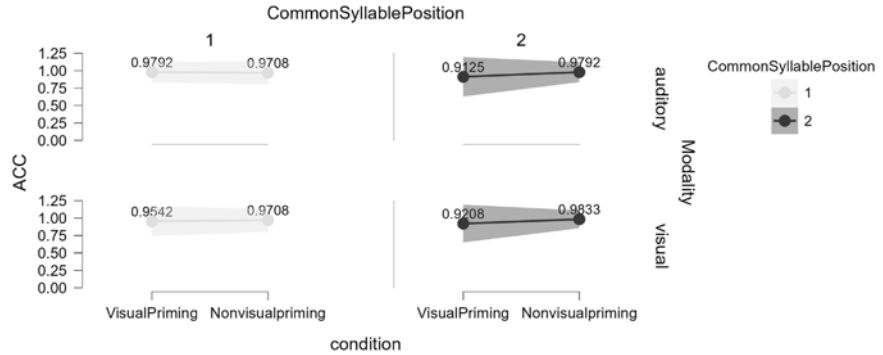



**Figure 2.** Participants' mean accuracy in each combination of the condition by the common syllable position by the modality; SD ribbon is shadowed.

was the first ($p$-Holm = 1) or the second ($p$-Holm = .082). Moreover, under both the orthographic ($p$-Holm = .191) and non-orthographic priming conditions ($p$-Holm = 1), accuracy did not differ significantly between first-syllable sharing and second-syllable sharing. The following linear mixed models were all built on the inverse reaction times recorded in those correctly-responded trials only.

The linear mixed model integrating both modalities showed that the 3-way interaction of the modality by the condition by the common syllable position was significant, $F(3, 877) = 5.207$, $p = .001$, and the 3-way interaction of the modality by the condition by the word class similarity was also significant, $F(3, 834) = 2.944$, $p = .032$.

After separating models by modalities, the model for the auditory session showed that participants' reaction times to targets were significantly modulated by the target word frequency, $F(1, 61) = 8.611$, $p = .005$ (Figure 5). In addition, the interaction of the condition by the word class similarity was not significant, $F(1, 164) = 0.673$, $p = .413$ (Table 3 & Figure 4), and the interaction of conditions by common syllable positions was marginally significant, $F(1, 151) = 3.494$, $p = .064$ (Table 3 & Figure 3).

The simple effect of conditions showed that participants' responses to the targets primed by shared 2nd orthography (e.g., 缺点 /que1 dian3/ *weakness* & 标点 /biao1 dian3/ *punctuation*) were significantly faster than their response speed to the targets primed by shared 2nd syllable without shared orthography (e.g., 缺点 /que1 dian3/ *weakness* & 经典 /jing1 dian3/ *classic*), $F(1, 124) = 21.764$, $p < .001$. In addition, participants' responses to the targets primed by shared 1st orthography (e.g., 目的 /mu4 di/ *goal* & 目录 /mu4 lu4/ *catalog*) were also significantly faster than to those targets primed by shared 1st syllable without shared orthography (e.g., 目的 /mu4 di/ *goal* & 木头 /mu4 tou/ *wood*), $F(1, 206) = 3.966$, $p = .048$.

The simple effect of common syllable positions showed that participants' responses to the targets primed by a shared 1st syllable (e.g., 目的 /mu4 di/ *goal* & 木头 /mu4 tou/ *wood*) were significantly faster than to those targets primed by a shared 2nd syllable (e.g., 缺点 /que1 dian3/ *weakness* & 经典 /jing1 dian3/ *classic*), $F(1, 79) = 5.375$, $p = .023$. However, participants' responses to the targets primed by shared 1st orthography (e.g., 目的 /mu4 di/ *goal* & 目录 /mu4 lu4/ *catalog*) did

https://doi.org/10.1017/S0142716426100812 Published online by Cambridge University Press

**Table 3.** The interactions of conditions by common syllable positions/word class similarity on the back-transformed reaction times in the (1) auditory session and (2) visual session

| | Shared orthography | Non-shared orthography | Differences |
|---|---|---|---|
| | (1) Auditory session | | |
| **1st syllable common** | 1296 ms (0.271) | 1383 ms (0.273) | 87 ms* |
| **2nd syllable common** | 1292 ms (0.267) | 1521 ms (0.273) | 229 ms*** |
| **Differences** | −4 ms | 138 ms* | |
| **Same word class** | 1262 ms (0.276) | 1440 ms (0.299) | 178 ms*** |
| **Different word class** | 1327 ms (0.255) | 1458 ms (0.243) | 131 ms** |
| **Differences** | 65 ms | 18 ms | |
| | (2) Visual session | | |
| **1st syllable common** | 898 ms (0.452) | 931 ms (0.452) | 33 ms |
| **2nd syllable common** | 955 ms (0.452) | 900 ms (0.450) | −55 ms |
| **Differences** | 57 ms | −31 ms | |
| **Same word class** | 899 ms (0.464) | 943 ms (0.498) | 44 ms |
| **Different word class** | 954 ms (0.442) | 889 ms (0.418) | −65 ms* |
| **Differences** | 55 ms | −54 ms | |

*$p$ <. 05, **$p$ <. 01, ***$p$ <. 001; SE in parentheses.

not differ significantly from their response speed to those targets primed by shared 2nd orthography (缺点 /que1 dian3/ *weakness* & 标点 /biao1 dian3/ *punctuation*), $F(1, 90) = 0.007$, $p = .933$.

To better illustrate the above simple effects, each participant's average inverse reaction time under the shared orthographic condition was back-transformed and then subtracted from that under the non-shared orthographic condition. This difference score served as an index of the individual participant's orthographic priming effect. The magnitude of each participant's orthographic priming effect was then ranked and plotted on a scale from smallest to largest in Figure 3. These steps were performed separately for each combination of modality (auditory, visual) and common syllable position (first and second).

The linear mixed model for the visual session showed significant interaction of the condition by the common syllable position, $F(1, 175) = 4.398$, $p = .037$ (Table 3 & Figure 3). The simple effect of conditions showed that participants' responses to the targets primed by shared 2nd orthography (e.g., 缺点 /que1 dian3/ *weakness* & 标点 /biao1 dian3/ *punctuation*) did not differ significantly from their response speed to those targets primed by shared 2nd syllable without shared orthography (e.g., 缺点 /que1 dian3/ *weakness* & 经典 /jing1 dian3/ *classic*), $F(1, 118) = 2.920$, $p = .090$. Meanwhile, participants' responses to target words primed by shared 1st orthography (e.g., 目的 /mu4 di/ *goal* & 目录 /mu4 lu4/ *catalog*) did not differ significantly from their response speed to those targets primed by shared 1st syllable without shared orthography (e.g., 目的 /mu4 di/ *goal* & 木头 /mu4 tou/ *wood*), $F(1, 326) = 1.495$, $p = .222$.

https://doi.org/10.1017/S0142716426100812 Published online by Cambridge University Press

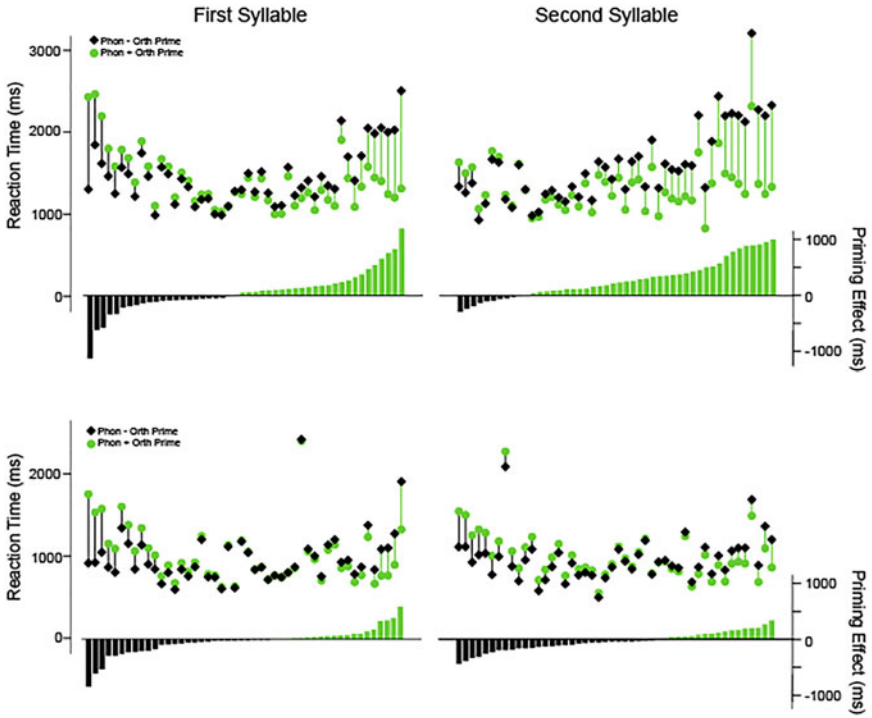



**Figure 3.** Each participant's orthographic priming effect, calculated as the back-transformed reaction time in the non-shared orthography condition (◆) minus that in the shared orthography condition (●), is displayed separately for each combination of modality and common syllable position. Top left: auditory session, first-syllable shared; top right: auditory session, second-syllable shared; bottom left: visual session, first-syllable shared; bottom right: visual session, second-syllable shared.

The simple effect of common syllable positions showed that participants' responses to the targets primed by a shared 1st syllable (e.g., 目的 /mu4 di/ *goal* & 木头 /mu4 tou/ *wood*) did not differ significantly from their response speed to those targets primed by a shared 2nd syllable (e.g., 缺点 /que1 dian3/ *weakness* & 经典 / jing1 dian3/ *classic*), $F(1, 83) = 0.932$, $p = .337$. In addition, participants' responses to the targets primed by a shared 1st orthography (e.g., 目的 /mu4 di/ *goal* & 目录 / mu4 lu4/ *catalog*) did not differ significantly from their response speed to those targets primed by shared 2nd orthography (缺点 /que1 dian3/ *weakness* & 标点 / biao1 dian3/ *punctuation*), $F(1, 88) = 2.544$, $p = .114$.

In addition to the significant interaction of conditions by common syllable positions, the interaction of the condition with the word class similarity was also significant, $F(1, 225) = 6.033$, $p = .015$ (Table 3 & Figure 4). The simple effect of conditions showed that participants' responses to the targets primed by a different word class with shared orthography were significantly slower than to those targets primed by a different word class without shared orthography, $F(1, 155) = 5.933$, $p = .016$. However, participants' responses to the targets primed by the same word class with shared orthography did not differ significantly from their response speed

https://doi.org/10.1017/S0142716426100812 Published online by Cambridge University Press




**Figure 4.** Back-transformed reaction times to the target primed by a word from the same word class (●) or from a different word class (▲) in each condition of each modality.

to those targets primed by the same word class without shared orthography, $F(1, 235) = 1.682$, $p = .196$.

The simple effect of the word class similarity showed that participants' responses to the targets primed by a different word class with shared orthography did not differ significantly from those targets primed by the same word class with shared orthography, $F(1, 96) = 2.279$, $p = .134$. In addition, participants' responses to the targets primed by a different word class without shared orthography did not differ significantly from their response speed to those targets primed by the same word class without shared orthography, $F(1, 101) = 2.309$, $p = .132$.

In the visual session, participants' response speed to targets was not significantly influenced by the log10 of target frequency, $F(1, 68) = 0.567$, $p = .454$ (Figure 5).

## Discussion

The present findings reveal a systematic dissociation between auditory and visual semantic processing in Chinese, demonstrating that orthographic representations are strategically recruited during spoken word recognition yet play a markedly different role in reading. Across both modalities, the semantic relatedness judgment task yielded high accuracy rates that were unaffected by orthographic overlap, confirming that participants engaged in genuine semantic evaluation rather than relying on superficial form-based strategies. The critical pattern emerging from the reaction time data is that orthographic priming facilitates auditory semantic access, particularly when the shared character occurs in the second syllable, where disambiguation demands peak, while exerting no comparable facilitatory influence and, under certain conditions, even interference, in visual processing. This cross-modal asymmetry, coupled with the selective modulation by target word frequency in audition and by word-class congruity in vision, indicates that orthography

https://doi.org/10.1017/S0142716426100812 Published online by Cambridge University Press

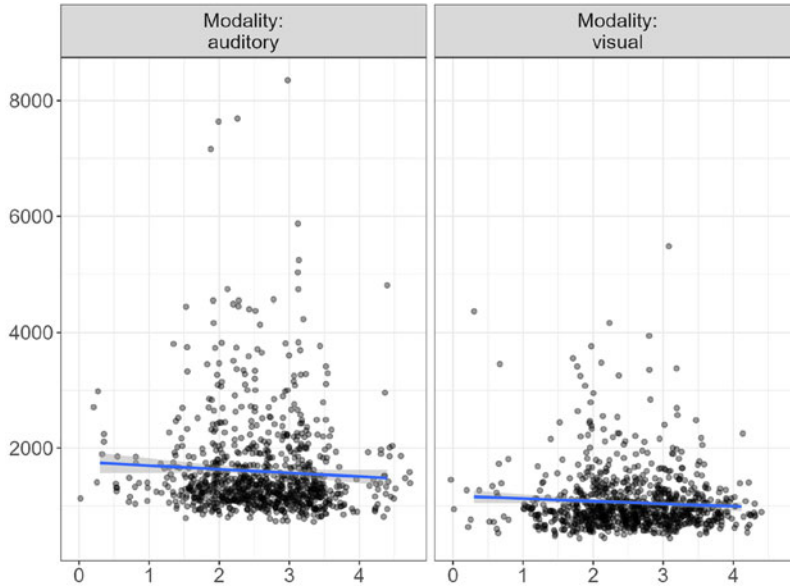



**Figure 5.** The correlation of back-transformed reaction times (y-axis) with the log10 of target word frequency (x-axis) in the auditory (left) and the visual session (right).

interacts with semantic processing through fundamentally distinct computational routes: an incremental, temporally constrained pathway in spoken word recognition that exploits orthographic codes to resolve phonological ambiguity, versus a parallel, direct mapping route in reading that renders additional orthographic overlap functionally redundant or potentially conflicting.

### *Interpreting accuracy and reaction time as complementary indices*

The generalized linear mixed model on accuracy rates revealed a significant main effect of condition and a significant condition × syllable-position interaction; however, follow-up simple-effects analyses showed that accuracy did not differ significantly between the shared-orthography and non-shared-orthography conditions within either the first-syllable or the second-syllable context (all $p$-Holm $> .05$). Moreover, modality did not significantly affect accuracy, nor did it interact with condition.

These accuracy patterns carry two important implications. First, the absence of a modality difference in accuracy confirms that the auditory and visual tasks were roughly equivalent in overall difficulty, thereby validating the cross-modal comparison of RTs. Had accuracy differed across modalities, any RT dissociation could have been attributed to differential task demands rather than to modality-specific lexical access pathways. Second, because accuracy was generally high and was not significantly modulated by orthographic overlap in the simple-effects analyses, the observed RT facilitation in the auditory modality cannot be attributed to a speed–accuracy trade-off. In other words, participants did not sacrifice accuracy for speed under the shared-orthography condition; rather, orthographic overlap genuinely accelerated the

https://doi.org/10.1017/S0142716426100812 Published online by Cambridge University Press

efficiency of semantic access without altering the outcome of the semantic decision. This dissociation, whereby a variable influences the latency but not the accuracy of a high-accuracy task, is common in semantic priming research and indicates that RT is the more sensitive dependent measure for capturing incremental processing dynamics (Baayen & Milin, 2010; Lachaud & Renaud, 2011). Thus, although the theoretical claims of this manuscript rely primarily on RT data, the accuracy results serve as converging evidence that the task was well-calibrated across modalities and that participants performed the semantic judgment correctly regardless of orthographic overlap.

### *Modality-specific orthographic priming*

The selective emergence of orthographic priming in auditory processing coheres with the architecture of interactive activation models (Grainger & Ferrand, 1996; Plaut et al., 1996), which posit bidirectional phonology–orthography connections that become functionally critical when sequential input parsing generates lexical ambiguity. In Chinese, where a single syllable activates numerous homophonic characters, the auditory signal unfolds incrementally, giving rise to a broad cohort of phonologically similar competitors (Marslen-Wilson, 1987; Marslen-Wilson & Tyler, 1997). Orthographic representations are then recruited as top-down constraints that progressively narrow this candidate set, operating as a compensatory disambiguation mechanism (Perfetti & Tan, 1998). Consequently, the auditory-specific orthographic priming reflects not a task-specific artifact but the fundamental computational demand of resolving lexical competition under temporally extended, sequential input.

By contrast, visual word recognition presents a fundamentally different computational landscape. Because both characters of a disyllabic compound are available simultaneously, readers can map orthographic form directly onto semantic content without relying on incremental disambiguation (Cheng & Shih, 1988; Pollatsek et al., 2000). The holistic presence of orthographic information renders additional overlap between prime and target functionally redundant, thereby accounting for the absence of orthographic priming in the visual modality.

It is important, however, to circumscribe the strength of our cross-modal conclusions. Because auditory and visual stimuli necessarily differ in their temporal structure, sequential acoustic unfolding versus simultaneous orthographic presentation, direct statistical comparison of the magnitude of priming across modalities is methodologically problematic. Our analytic strategy of fitting separate linear mixed models for each modality, while statistically justified by the significant three-way interaction involving modality, condition, and common syllable position, precludes strong parametric claims about the difference in effect size between the two modalities. Consequently, the present data should be interpreted as consistent with modality-specific lexical access pathways rather than as definitive proof of representational modality independence. The auditory-preferential effect may, in part, reflect the differential temporal demands of sequential versus parallel input, and future research employing time-sensitive neurophysiological measures (e.g., EEG or MEG) will be needed to adjudicate whether the absence of visual priming stems from representational architecture, processing stage, or purely temporal factors.

https://doi.org/10.1017/S0142716426100812 Published online by Cambridge University Press

### *Common syllable position effects*

In the auditory modality, orthographic facilitation was numerically larger and more robust when the shared character occurred in the second syllable (229 ms; e.g., 缺点 /que1 dian3/ *weakness* & 标点 /biao1 dian3/ *punctuation*) than when it occurred in the first syllable (87 ms; e.g., 目的 /mu4 di/ *goal* & 目录 /mu4 lu4/ *catalog*), even though both achieved statistical significance. This positional gradient is readily interpretable within the cohort model of spoken word recognition (Marslen-Wilson, 1987; Marslen-Wilson & Welsh, 1978; Norris, 1994): upon hearing the first syllable, listeners activate a broad cohort of candidates that share that phonological onset, and in Chinese this initial cohort is exceptionally large due to high syllable-level homophony (Treiman et al., 1981; Yin & Rohsenow, 1994). When the second syllable arrives, it provides critical diagnostic information that sharply narrows the candidate set, and at this temporally precise "disambiguation window," the co-activation of a shared orthographic form supplies additional stabilizing constraint, thereby yielding stronger facilitation. This pattern is further supported by prior ERP and behavioral evidence that later-arriving phonological and orthographic information exerts greater influence on lexical selection in Chinese spoken word recognition (Lü et al., 2004; Yin et al., 2011). The fact that first-syllable overlap also produced significant facilitation, albeit smaller in magnitude, indicates that orthographic codes begin to co-activate early in the auditory signal, consistent with the continuous mapping assumption of connectionist models (Plaut et al., 1996), yet their facilitatory impact is necessarily diluted when the cohort remains large and ambiguous.

By contrast, the visual modality shows no reliable positional asymmetry because both characters are processed in parallel, eliminating the temporal distinction between "early" and "late" syllabic information. With orthographic forms available holistically, semantic access bypasses the need for incremental disambiguation, thereby eliminating positional effects. This dissociation between modalities underscores a fundamental processing-stage difference: auditory recognition relies on serial decoding in which orthographic priming is dynamically coupled to the temporal structure of the input, whereas visual recognition benefits from parallel access that permits direct orthographic-to-semantic mapping (Whitney, 1998).

### *Word-class similarity*

The interaction between orthographic priming and word-class similarity further highlights modality-specific mechanisms. In the auditory modality, orthographic facilitation occurred regardless of whether the prime and target belonged to the same or different word classes (e.g., noun-noun pairs vs. noun-verb pairs). This pattern suggests that orthographic co-activation in spoken word recognition is relatively automatic and does not depend on the convergence of syntactic–semantic category information. It is consistent with the view that phonological and orthographic nodes interact pre-lexically or at an early lexical stage, before full semantic integration (Fodor, 1983; Pulvermüller, 1999).

In the visual modality, by contrast, a significant condition × word-class interaction emerged: when word classes differed, shared orthography actually slowed

https://doi.org/10.1017/S0142716426100812 Published online by Cambridge University Press

responses relative to non-shared orthography. We interpret this as a post-lexical response-competition effect. When two words are orthographically dissimilar, belong to different word classes, and are semantically unrelated, the orthographic, syntactic, and semantic cues all converge on a "no" response, allowing a rapid and unambiguous decision. The critical contrast arises when two words share orthographic form yet remain semantically unrelated and cross-class. Here, visual overlap between prime and target generates perceptual fluency and form-based familiarity, which transiently primes a "yes" response. However, as the words belong to different word classes and are semantically unrelated, the correct judgment is "no." The result is a conflict between an orthographically driven "yes" bias and a semantically mandated "no" response. Resolving this competition requires engaging inhibitory control mechanisms to suppress the misleading orthographic cue (Posner & Petersen, 1990; van Orden et al., 1990), thereby consuming additional cognitive resources and producing longer reaction times in the shared than in the non-shared orthography condition. This pattern mirrors findings in alphabetic languages, where orthographic-linguistic mismatches similarly recruit cognitive control networks (van Orden et al., 1990).

### *Word frequency effects*

The significant word frequency effect in auditory but not visual processing further corroborates modality-specific pathways. In auditory word recognition, high-frequency targets are processed more efficiently because their stronger lexical representations facilitate faster resolution of cohort competition when the unfolding input reaches the critical disambiguation point (Harley, 2013). Because the auditory signal unfolds sequentially, frequency amplifies the efficiency with which orthographic information disambiguates the incoming signal. This pattern is consistent with connectionist accounts in which frequency effects emerge more robustly under sequential than simultaneous input conditions (Plaut et al., 1996). In visual processing, by contrast, orthographic forms are fully and simultaneously available, enabling direct semantic access that does not rely on temporally driven lexical competition. Frequency therefore exerts little influence on visual word recognition (Cai & Brysbaert, 2010). Taken together, the dissociation indicates that frequency effects emerge only when lexical access is temporally constrained: in audition, incremental input creates a window of competition in which high-frequency words are resolved faster; in vision, by contrast, parallel access eliminates this temporal bottleneck, rendering frequency effects redundant.

### *Theoretical integration and innovation*

While prior work has documented orthographic effects in Chinese (e.g., Qu & Damian, 2016; Zou et al., 2012), this study advances the field by providing a systematic cross-modal comparison within a semantic judgment paradigm that manipulates the orthographic overlap and the common syllable position and controls for confounds such as phonology, word frequency, cohort density, visual–spatial variables, and word-class similarity. This design offers a purer behavioral index of orthographic priming than prior lexical-decision studies (e.g., Zhou &

https://doi.org/10.1017/S0142716426100812 Published online by Cambridge University Press

Marslen-Wilson, 1995; Zou et al., 2012), allowing us to attribute the auditory facilitation more confidently to orthographic identity rather than to phonetic or lexical repetition artifacts.

Our findings refine interactive activation models by incorporating cross-linguistic evidence that underscores script-specificity in phonology–orthography interactions. In alphabetic languages, orthographic effects in audition are typically explained via systematic grapheme–phoneme correspondence rules (Ziegler & Ferrand, 1998; Ziegler et al., 2008). The present results demonstrate that in Chinese, which is a nonalphabetic system with opaque sound–spelling mappings, orthographic mediation in auditory processing operates not through spelling-to-sound translation but through homophone disambiguation and semantic radical activation. In addition, by revealing that orthographic codes are strategically exploited to resolve phonological uncertainty, these findings question phylogenetic and ontogenetic assumptions of direct sound-to-meaning access (MacWhinney & O'Grady, 2015), suggesting instead that literacy reshapes spoken language processing to exploit orthographic information in a manner that is both modality- and script-specific.

The absence of visual orthographic priming, however, cannot be reductively attributed to "direct semantic access." First, multimodal representation theories posit that spoken word recognition obligatorily recruits orthographic codes in literate individuals due to literacy-induced neural reorganization (Dehaene et al., 2010). In Chinese, where homophony is pervasive, this orthographic activation functions as a compensatory disambiguation mechanism. This is less critical in visual processing, where characters directly cue semantics (Kahneman, 1973). Second, processing-stage theories suggest that orthographic effects in audition arise relatively early to reduce candidate sets, whereas in vision they may emerge post-lexically and induce response conflict rather than facilitation (Coltheart et al., 2001). This is supported by our results that when the prime and the target came from different word classes, visual orthographic overlap produced interference rather than facilitation, and also by prior findings that when Chinese primes and targets were already homophonic (e.g., 诚/cheng2/ *honest* & 城/cheng2/ *city*), additional orthographic similarity did not yield extra facilitation (Cheng & Shih, 1988; Pollatsek et al., 2000).

## Conclusion and limitations

In summary, the present study provides robust evidence that orthographic priming facilitates Chinese auditory semantic processing, particularly when orthographic overlap occurs in the second syllable where disambiguation demands are highest. This facilitation is accompanied by frequency-sensitive modulation and is immune to word-class manipulation, suggesting early automatic co-activation of orthographic and phonological codes. In visual processing, by contrast, orthographic overlap does not facilitate semantic judgment and may even induce interference under cross-class pairs, a pattern consistent with direct orthographic-to-semantic access and post-lexical response competition. These results cohere with interactive

https://doi.org/10.1017/S0142716426100812 Published online by Cambridge University Press

activation models while emphasizing the unique role of Chinese characters in bridging sound and meaning.

Several limitations exist. First, as noted above, the inherent temporal disparity between auditory and visual presentation formats limits the strength of inferences about modality-specific representational architectures. Future studies should employ temporally aligned designs or online neurophysiological recordings to track the time course of orthographic activation in both modalities. Second, the small trial number per condition (10 trials) may limit statistical power, though linear mixed models mitigated this issue by accounting for random effects. Future studies should expand stimuli to enhance generalizability.

**Acknowledgments.** The research was supported by the 2021 Guangdong Provincial General Universities Characteristic Innovative Project entitled "A Psycholinguistic Study of Chinese Adverbs" 2021年度广东省普通高校特色创新类项目"汉语副词的心理语言学研究"(2021WTSCX025). We are deeply grateful to Prof. Annabel Chen, Prof. Suzy Styles, and Prof. Chiao-Yi Wu for their intellectual guidance in the earliest phase of this work.

## Notes

**1** Chinese characters are monosyllabic. In pinyin transcription (the Romanization system for Chinese), a number following a syllable indicates the lexical tone: 1 = high-level tone, 2 = rising tone, 3 = falling-rising tone, and 4 = falling tone. Notably, the null/neutral tone is not marked by a number. For example, "/ye4/" represents the syllable "ye" spoken with the falling tone, while "/ye/" without a number indicates the null/neutral tone.

**2** Cohort density refers to the number of existing disyllabic words that share the same first syllable with a given disyllabic word. Based on the cohort theory (Marslen-Wilson, 1987; Marslen-Wilson & Welsh, 1978) and the shortlist model (Norris, 1994), cohort density is demonstrated to affect participants' reaction time to spoken words (e.g., Magnuson et al., 2007). For example, the cohort members of 目的 (/mu4 di/ *goal*) include 牧羊 (/mu4 yang2/ *shepherd*), 目前 (/mu4 qian2/ *at present*), 木瓜 (/mu4 gua1/ *papaya*), 慕名 (/mu4 ming2/ *to admire somebody's reputation*), etc. By contrast, there are only three cohort members for 笼罩 (/long3 zhao4/ *to envelop*), namely, 垄断 (/long3 duan4/ *enjoy market dominance*), 笼络 (/long3 luo4/ *to win somebody over*), and 笼统 (/long3 tong3/ *general*). Thus, a longer time may be needed to disambiguate the word with a denser cohort, e.g., 目的 (/mu4 di/ *goal*) > 笼罩 (/long3 zhao4/ *to envelop*), as listeners would have higher uncertainty about what word they are hearing by the end of the first syllable and thus have to wait for more phonological information to be unfolded.

**3** The homophonic syllable appeared in a consistent position within each pair of words, as inconsistent positions may inhibit the effect of orthography (Ding et al., 2004), such as a counter-example: 照 (/zhao4/ *shine*) is the first syllable of 照耀 (/zhao4 yao4/ *illuminate*) but the second syllable of 按照 (/an4 zhao4/ *according to*).

## References

**Baayen, R. H., & Milin, P.** (2010). Analyzing reaction times. *International Journal of Psychological Research*, **3**(2), 12–28.

**Cai, Q., & Brysbaert, M.** (2010). SUBTLEX-CH: Chinese word and character frequencies based on film subtitles. *PLoS One*, **5**(6), e10729. https://doi.org/10.1371/journal.pone.0010729.

**Chao, Y. R.** (1976). *Aspects of Chinese sociolinguistics: Essays* (Vol. **9**). Stanford University Press.

**Chen, W., Chao, P., Chang, Y., Hsu, C., & Lee, C.** (2016). Effects of orthographic consistency and homophone density on Chinese spoken word recognition. *Brain and Language*, **157–158**, 51–62. https://doi.org/10.1016/j.bandl.2016.08.002.

https://doi.org/10.1017/S0142716426100812 Published online by Cambridge University Press

**Cheng, C.-M., & Shih, S.** (1988). The Nature of Lexical Access in Chinese: Evidence from Experiments on Visual and Phonological Priming in Lexical Judgment. In I. M. Liu (Ed.), *Cognitive aspects of the Chinese language* (Vol. **1**, pp. 1–14). Asian Research Service.

**Chéreau, C., Gaskell, M. G., & Dumay, N.** (2007). Reading spoken words: Orthographic effects in auditory priming. *Cognition*, **102**(3), 341–360. https://doi.org/10.1016/j.cognition.2006.01.001.

**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. https://doi.org/10.1037/0033-295X.108.1.204.

**Dehaene, S., Pegado, F., Braga, L. W., Ventura, P., Nunes Filho, G., Jobert, A., & Cohen, L.** (2010). How learning to read changes the cortical networks for vision and language. *Science*, **330**(6009), 1359–1364. https://doi.org/10.1126/science.1194140.

**Ding, G., Peng, D., & Taft, M.** (2004). The nature of the mental representation of radicals in Chinese: A priming study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **30**(2), 530–539. https://doi.org/10.1037/0278-7393.30.2.530.

**Dumay, N., Benraïss, A., Barriol, B., Colin, C., Radeau, M., & Besson, M.** (2001). Behavioral and electrophysiological study of phonological priming between bisyllabic spoken words. *Journal of Cognitive Neuroscience*, **13**(1), 121–143.

**Feng, G., Miller, K., Shu, H., & Zhang, H.** (2001). Rowed to recovery: The use of phonological and orthographic information in reading Chinese and English. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **27**(4), 1079–1100. https://doi.org/10.1037/0278-7393.27.4.1079.

**Fodor, J. A.** (1983). *The modularity of mind*. MIT Press.

**Frauenfelder, U. H., Segui, J., & Dijkstra, T.** (1990). Lexical effects in phonemic processing: Facilitatory or inhibitory? *Journal of Experimental Psychology: Human Perception and Performance*, **16**(1), 77–91.

**Frith, U.** (1998). Literally changing the brain. *Brain*, **121**(6), 1011–1012.

**Grainger, J., & Ferrand, L.** (1996). Masked orthographic and phonological priming in visual word recognition and naming: Cross-task comparisons. *Journal of Memory and Language*, **35**(5), 623–647. https://doi.org/10.1006/jmla.1996.0033.

**Harley, T. A.** (2013). *The psychology of language: From data to theory* (4th ed.). Psychology Press.

**Kahneman, D.** (1973). *Attention and effort*. Prentice-Hall.

**Lachaud, C. M., & Renaud, O.** (2011). A tutorial for analyzing human reaction times: How to filter data, manage missing values, and choose a statistical model. *Applied Psycholinguistics*, **32**(2), 389–416. https://doi.org/10.1017/S0142716410000457.

**Lü, Y., Shen, D. L., Du, Y. C., & Han, Z. Y.** (2004). 听觉呈现条件下汉语双字词语义和语音启动的ERP研究 [An ERP study on semantic and phonological priming of two-character words under auditory presentation condition]. *心理科学 (Psychological Science)*, **27**(1), 8–12.

**MacWhinney, B., & O'Grady, W.** (2015). *The handbook of language emergence*. Wiley-Blackwell.

**Magnuson, J. S., Dixon, J. A., Tanenhaus, M. K., & Aslin, R. N.** (2007). The dynamics of lexical competition during spoken word recognition. *Cognitive Science*, **31**(1), 133–156. https://doi.org/10.1080/03640210709336987.

**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., & Tyler, L. K.** (1997). Dissociating types of mental computation. *Nature*, **387**(6633), 592–594. https://doi.org/10.1038/42436.

**Marslen-Wilson, W. D., & Welsh, A.** (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology*, **10**(1), 29–63.

**Miller, K. M., & Swick, D.** (2003). Orthography influences the perception of speech in alexic patients. *Journal of Cognitive Neuroscience*, **15**(7), 981–990. https://doi.org/10.1162/089892903770007356.

**Muneaux, M., & Ziegler, J.** (2004). Locus of orthographic effects in spoken word recognition: Novel insights from the neighbour generation task. *Language and Cognitive Processes*, **19**(5), 641–660. https://doi.org/10.1080/01690960444000052.

**Norris, D.** (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*, **52**(3), 189–234.

**Oldfield, R. C.** (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, **9**(1), 97–113.

https://doi.org/10.1017/S0142716426100812 Published online by Cambridge University Press

**Paap, K. R., & Noel, R. W.** (1991). Dual-route models of print and sound: Still a good horse race. *Psychological Research*, **53**(1), 13–24.

**Pattamadilok, C., Kolinsky, R., Luksaneeyanawin, S., & Morais, J.** (2008). Orthographic congruency effects in the suprasegmental domain: Evidence from Thai. *The Quarterly Journal of Experimental Psychology*, **61**(10), 1515–1537. https://doi.org/10.1080/17470210701638931

**Pattamadilok, C., Morais, J., Ventura, P., & Kolinsky, R.** (2007). The locus of the orthographic consistency effect in auditory word recognition: Further evidence from French. *Language and Cognitive Processes*, **22**(1), 1–27. https://doi.org/10.1080/01690960500268261.

**Pattamadilok, C., Perre, L., Dufau, S., & Ziegler, J. C.** (2009). On-line orthographic influences on spoken language in a semantic task. *Journal of Cognitive Neuroscience*, **21**(1), 169–179. https://doi.org/10.1162/jocn.2008.21002.

**Perfetti, C. A., & Tan, L. H.** (1998). The time course of graphic, phonological, and semantic activation in Chinese character identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **24**(1), 101–118. https://doi.org/10.1037/0278-7393.24.1.101.

**Perre, L., Bertrand, D., & Ziegler, J.** (2011). Literacy affects spoken language in a nonlinguistic task: An ERP study. *Frontiers in Psychology*, **2**, 274. https://doi.org/10.3389/fpsyg.2011.00274.

**Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K.** (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review*, **103**(1), 56–115. https://doi.org/10.1037/0033-295X.103.1.56.

**Pollatsek, A., Tan, L. H., & Rayner, K.** (2000). The role of phonological codes in integrating information across saccadic eye movements in Chinese character identification. *Journal of Experimental Psychology: Human Perception and Performance*, **26**(2), 607–633.

**Posner, M. I., & Petersen, S. E.** (1990). The attention system of the human brain. *Annual Review of Neuroscience*, **13**(1), 25–42. https://doi.org/10.1146/annurev.ne.13.030190.000325.

**Pulvermüller, F.** (1999). Words in the brain's language. *Behavioral and Brain Sciences*, **22**(2), 253–279. https://doi.org/10.1017/S0140525X9900182X.

**Qu, Q., & Damian, M. F.** (2016). Orthographic effects in spoken word recognition: Evidence from Chinese. *Psychonomic Bulletin & Review*, **23**(6), 1883–1889. https://doi.org/10.3758/s13423-016-1164-9.

**Rodd, J. M., Davis, M. H., & Johnsrude, I. S.** (2005). The neural mechanisms of speech comprehension: fMRI studies of semantic ambiguity. *Cerebral Cortex*, **15**(8), 1261–1269. https://doi.org/10.1093/cercor/bhi009.

**Schneider, W., Eschman, A., & Zuccolotto, A.** (2002). *E-Prime reference guide*. Psychology Software Tools.

**Seidenberg, M. S., & Tanenhaus, M. K.** (1979). Orthographic effects on rhyme monitoring. *Journal of Experimental Psychology: Human Learning and Memory*, **5**(6), 546–554.

**Slowiaczek, L. M., & Hamburger, M.** (1992). Prelexical facilitation and lexical interference in auditory word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **18**(6), 1239–1253.

**Slowiaczek, L. M., Soltano, E. G., Wieting, S. J., & Bishop, K. L.** (2003). An investigation of phonology and orthography in spoken-word recognition. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, **56**(2), 233–262. https://doi.org/10.1080/02724980244000332.

**Taft, M., Castles, A., Davis, C., Lazendic, G., & Nguyen-Hoan, M.** (2008). Automatic activation of orthography in spoken word recognition: Pseudohomograph priming. *Journal of Memory and Language*, **58**(2), 366–379. https://doi.org/10.1016/j.jml.2007.05.002.

**Treiman, R., Baron, J., & Luk, K.** (1981). Speech recoding in silent reading: A comparison of Chinese and English. *Journal of Chinese Linguistics*, **9**(1), 116–125.

**van Orden, G. C., Pennington, B. F., & Stone, G. O.** (1990). Word identification in reading and the promise of subsymbolic psycholinguistics. *Psychological Review*, **97**(4), 488–522. https://doi.org/10.1037/0033-295X.97.4.488.

**Ventura, P., Morais, J., Pattamadilok, C., & Kolinsky, R.** (2004). The locus of the orthographic consistency effect in auditory word recognition. *Language and Cognitive Processes*, **19**(1), 57–95. https://doi.org/10.1080/01690960344000134.

**Wang, H., Chang, B. R., & Li, Y. S.** (1986). *现代汉语频率词典 [Frequency dictionary of modern Chinese]*. Beijing Language and Culture University Press.

**Wang, H., Li, Y., Ning, Q., & Zhang, J.** (2012). The role of homophone density in modulating whole-character effects during Chinese spoken word recognition. *Acta Psychological Sinica*, **44**(5), 567–578. https://doi.org/10.3724/SP.J.1041.2012.00567.

https://doi.org/10.1017/S0142716426100812 Published online by Cambridge University Press

**Whelan, R.** (2008). Effective analysis of reaction time data. *The Psychological Record*, **58**(3), 475–482.
**Whitney, P.** (1998). *The psychology of language*. Houghton Mifflin.
**Yin, B., & Rohsenow, J. S.** (1994). *Modern Chinese characters*. Sinolingua.
**Yin, C., Wang, J., & Zhang, J. J.** (2011). 汉语言语产生中语音、字形启动的位置效应 [The position effect of phonological and orthographic priming in Chinese spoken word production]. *心理学报 (Acta Psychological Sinica)*, **43**(9), 1002–1012.
**Zhang, H.** (2010). *汉语近义词词典 [A dictionary of Chinese near-synonyms]*. Beijing: The Commercial Press.
**Zhao, X., & Li, Y.** (2009). *商务馆学汉语近义词词典 [The Commercial Press dictionary of Chinese near-synonyms for learners]*. The Commercial Press.
**Zhou, X.** (1978). 汉字音形映射的任意性研究 [The arbitrariness of sound-spelling mapping in Chinese characters]. *语言研究 (Linguistic Research)*, **2**(1), 15–30.
**Zhou, X., & Marslen-Wilson, W.** (1995). Morphological structure in the Chinese mental lexicon. *Language and Cognitive Processes*, **10**(6), 545–600. https://doi.org/10.1080/01690969508407114.
**Zhu, X.** (1988). 现代汉字声旁表音功能的动态 [Analysis of Cuing Function of Phonetic Components in Modern Chinese]. In X. Yuan (Ed.), *中国语言文字学学术研讨会论文集 [Proceedings of the symposium on the Chinese language and characters]* (pp. 260–288). Guang Ming Daily Press.
**Ziegler, J. C., & Ferrand, L.** (1998). Orthography shapes the perception of speech: The consistency effect in auditory word recognition. *Psychonomic Bulletin & Review*, **5**(4), 683–689. https://doi.org/10.3758/BF03208845.
**Ziegler, J. C., Muneaux, M., & Grainger, J.** (2003). Neighborhood effects in auditory word recognition: Phonological competition and orthographic facilitation. *Journal of Memory and Language*, **48**(4), 779–793. https://doi.org/10.1016/S0749-596X(03)00023-7.
**Ziegler, J. C., Petrova, A., & Ferrand, L.** (2008). Feedback consistency effects in visual and auditory word recognition: Where do we stand after more than a decade? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **34**(3), 643–661. https://doi.org/10.1037/0278-7393.34.3.643.
**Zou, L., Desroches, A. S., Liu, Y., Xia, Z., & Shu, H.** (2012). Orthographic facilitation in Chinese spoken word recognition: An ERP study. *Brain and Language*, **123**(3), 164–173. https://doi.org/10.1016/j.bandl.2012.09.006.
**Zou, L., Packard, J. L., Xia, Z., Liu, Y., & Shu, H.** (2015). Neural correlates of morphological processing: Evidence from Chinese. *Frontiers in Human Neuroscience*, **9**, 555. https://doi.org/10.3389/fnhum.2015.00555.

**Cite this article:** Huang, T. & Liu, H. (2026). Modality-specific orthographic priming in Chinese semantic processing. *Applied Psycholinguistics*. https://doi.org/10.1017/S0142716426100812

https://doi.org/10.1017/S0142716426100812 Published online by Cambridge University Press