Orthographic properties of distractors do influence phonological Stroop effects: Evidence from Japanese Romaji distractors

Masahiro Yoshihara[1,2] • Mariko Nakayama[3] • Rinus G. Verdonschot[4] • Yasushi Hino[2] • Stephen J. Lupker[5]

Abstract

In attempting to understand mental processes, it is important to use a task that appropriately reflects the underlying processes being investigated. Recently, Verdonschot and Kinoshita (*Memory & Cognition, 46,* 410–425, 2018) proposed that a variant of the Stroop task—the "phonological Stroop task"—might be a suitable tool for investigating speech production. The major advantage of this task is that the task is apparently not affected by the orthographic properties of the stimuli, unlike other, commonly used, tasks (e.g., associative-cuing and word-reading tasks). The viability of this proposal was examined in the present experiments by manipulating the script types of Japanese distractors. For Romaji distractors (e.g., "kushi"), color-naming responses were faster when the initial phoneme was shared between the color name and the distractor than when the initial phonemes were different, thereby showing a phoneme-based phonological Stroop effect (Experiment 1). In contrast, no such effect was observed when the same distractors were presented in Katakana (e.g., "くし"), replicating Verdonschot and Kinoshita's original results (Experiment 2). A phoneme-based effect was again found when the Katakana distractors used in Verdonschot and Kinoshita's original study were transcribed and presented in Romaji (Experiment 3). Because the observation of a phonemic effect *directly depended on the orthographic properties of the distractor stimuli*, we conclude that the phonological Stroop task is also susceptible to orthographic influences.

Keywords Phonological Stroop tasks · Phoneme · Romaji · Japanese language

The main goal of speech production research is to understand the processes involved in the speech production system. One subgoal is to clarify how the initial phonological unit, which is the first selectable functional unit produced in the phonological encoding stage, varies across languages (e.g., O'Séaghdha, 2015; O'Séaghdha, Chen, & Chen, 2010; Roelofs, 2015). According to theory, phonological units are spelled out in parallel and then are incrementally inserted into a metrical frame indexing the number of syllables and (for some languages) the stress pattern (e.g., Levelt, Roelofs, & Meyer, 1999). The phonological unit in Indo-European languages such as English and Dutch is assumed to be the phoneme, whereas research has suggested that the counterpart in Mandarin Chinese is the (atonal) syllable (e.g., O'Séaghdha, 2015; O'Séaghdha et al., 2010; Roelofs, 2015).

In recent speech production research, a great deal of effort has been devoted to examining the nature of the phonological unit (e.g., Chen, O'Séaghdha, & Chen, 2016; Han & Verdonschot, 2019; Kureta, Fushimi, & Tatsumi, 2006; Nakayama, Kinoshita, & Verdonschot, 2016; O'Séaghdha et al., 2010; Wong, Chiu, Wang, Wong, & Chen, 2019; Yoshihara, Nakayama, Verdonschot, & Hino, 2017). The main goal of the present research is, however, not to examine the phonological unit itself. Rather, the goal is to evaluate the nature of an experimental task used to investigate the phonological unit. Specifically, we examined what has been referred to as "the phonological Stroop task," a task that has been recently suggested as being well suited for research into the nature of the phonological unit based on the idea that it *is not*

✉ Masahiro Yoshihara
masahiro.yoshihara@gmail.com

1 Present address: Center for Japanese-Language Testing, The Japan Foundation, Shinjuku-ku, Tokyo, Japan

2 Faculty of Letters, Arts, and Sciences, Waseda University, Shinjuku-ku, Tokyo, Japan

3 Graduate School of International Cultural Studies, Tohoku University, Sendai, Miyagi, Japan

4 Department of Oral and Maxillofacial Radiology, Institute of Biomedical & Health Sciences, Hiroshima University, Hiroshima, Hiroshima, Japan

5 Department of Psychology, University of Western Ontario, London, ON, Canada

affected by the orthography of the stimuli (Verdonschot & Kinoshita, 2018). To explain the rationale for the present research, we will begin by briefly reviewing the typical experimental tasks used in speech production research.

Research investigating the phonological unit has typically involved one of two experimental tasks. One task is a word-reading task coupled with the masked priming paradigm (e.g., Nakayama et al., 2016; Verdonschot et al., 2011; You, Zhang, & Verdonschot, 2012). Participants in this task are instructed to read aloud a visually presented target preceded by the brief (e.g., 50 ms) presentation of a prime stimulus. The prime either has the same initial phonological segment(s) as the target (e.g., <u>belly</u>–*BREAK*) or is phonologically unrelated to the target (e.g., *merry*–*BREAK*). The speed and accuracy of reading targets aloud are the dependent measures in this task.

The other task is the associative-cuing task coupled with the form preparation (a.k.a., implicit priming) paradigm (e.g., Kureta et al., 2006; Meyer, 1990, 1991; O'Séaghdha et al., 2010). In this task, participants are first instructed to remember small sets of associatively related word pairs (e.g., *night–day, tint–dye, bread–dough, wet–dew*). After doing so, participants are presented with a prompt word (e.g., *night*), to which they are asked to produce the paired response word (i.e., *day*). Response words in a set either have the same initial phonological segment(s) (e.g., <u>d</u>ay, <u>d</u>ye, <u>d</u>ough, <u>d</u>ew) or do not (e.g., *day*, *pea*, *rye*, *sow*). The speed and accuracy of producing the response words are the dependent measures in the task.

In both tasks, the minimum size of the shared phonological segment(s) needed to observe a facilitation effect is thought to index the phonological unit size in the language being investigated. Typically, in each language investigated, the results in the two tasks have converged in suggesting the identity of the unit. For Dutch and English, significant facilitation due to phonemic (segmental) overlap has been observed in both tasks (e.g., masked priming: Forster & Davis, 1991; form preparation: Meyer, 1990, 1991). For Mandarin Chinese, facilitation has been observed due to syllabic overlap, but not due to phonemic overlap in both tasks (e.g., masked priming: You et al., 2012; form preparation: O'Séaghdha et al., 2010).

One of the important issues researchers have been interested in is whether the results of these tasks are affected by orthographic properties of a to-be-produced word (e.g., Alario, Perre, Castel, & Ziegler, 2007; Bi, Wei, Janssen, & Han, 2009; Damian & Bowers, 2003; Kureta, Fushimi, Sakuma, & Tatsumi, 2015; Li & Wang, 2017; Roelofs, 2006). For example, Damian and Bowers (2003), using English stimuli, reported that the form-preparation effect was not observed in the associative-cuing task when the response words had the same initial phoneme written with different letters (e.g., <u>c</u>amel, <u>k</u>ayak, <u>k</u>idney). This result indicates that the form-preparation effect depends on the orthographic nature of stimuli. It should be noted, however, that Damian

and Bowers's result has not been replicated (e.g., Alario et al., 2007; Bi et al., 2009; Roelofs, 2006).

Orthographic influences on speech production have also been reported in Japanese language experiments. Japanese is unique in that it uses multiple types of scripts simultaneously (Kanji, Kana, and, to a lesser extent, Romaji). Kanji, originally imported from Chinese, is a logographic script (e.g., 安以宇衣於). As most Kanji characters represent meaning (e.g., 赤 /a.ka/ "red"), they are basically used for (stems of) content words such as nouns, adverbs, and verbs in Japanese sentences. On the other hand, the two types of Kanas, Hiragana and Katakana, which were developed from Kanji, are syllabaries. In general, Hiragana is more cursive (e.g., あいうえお) and is used for grammatical elements such as adjective/verb inflections and grammatical particles, whereas Katakana is more angular (e.g., アイウエオ) and is mainly used for loan words. In contrast to Kanji and Kana, Romaji involves writing Japanese using the Roman alphabet (e.g., aiueo), and is mainly used to transcribe Japanese words for non-Japanese readers (although many Japanese people use Romaji to input Japanese words into smartphones and computers). Although most Japanese words tend to be written in a particular script, any Kanji word can be transcribed into either Kana or Romaji (though the reverse is not always true; for example, テーブル /te.R.bu.ru/ "table" cannot be written in Kanji). In addition, any Kana word can be transcribed into Romaji, and vice versa.

This characteristic of the Japanese language may have something to do with producing an apparent inconsistency in Japanese speech production research. In general, most research investigating the phonological unit in Japanese has suggested that the phonological unit is the mora (e.g., Kureta et al., 2006; Verdonschot et al., 2011) which is a temporal, syllabic-type unit of a roughly constant duration (e.g., Warner & Arai, 2001). However, the data in the literature are still somewhat inconclusive on this point.

In the masked priming word-reading task, mora-based facilitation, but no phoneme-based facilitation, has been observed when the stimuli are presented in Kana (Verdonschot et al., 2011). That is, a masked priming effect was observed when prime–target pairs had the same initial mora (e.g., スミ /<u>su</u>.mi/ "corner"–すし /<u>su</u>.si/ "sushi"), but not when they had the same initial phoneme (e.g., せん /<u>se</u>.N/ "line"–すし /<u>su</u>.si/). This data pattern was also observed when Kana stimuli were transcribed into Romaji (Verdonschot et al., 2011, Experiments 2 and 3).

When the stimuli were Kanji compound words, however, a different data pattern emerged: responses were not always facilitated when the prime–target pairs shared their initial mora (Yoshihara et al., 2017). Specifically, mora-based facilitation occurred only when the shared mora corresponded to *the whole sound* of the prime–target pairs' initial Kanji characters (i.e., there was a significant priming effect for a pair like 化石

/**ka**-se.ki/ "fossil"–火力 /**ka**-rjo.ku/ "heating power," but there was no priming for a pair like 確保 /**ka.ku**-ho/ "security"–火力 /**ka**-ryo.ku/, the bold morae represent the pronunciations of the first Kanji characters). Interestingly, however, when the Kanji compound words were transcribed into Kana, a standard mora-based priming effect emerged in both cases. That is, the primes かせき /**ka**.se.ki/ and かくほ /**ka**.ku.ho/ both facilitated the reading of the target カリョク /**ka**.ryo.ku/. (Note that the bold morae correspond to the pronunciations of the first Kana characters here.) Because the size of phonological overlap that produced the masked priming effects matched the size of phonology carried by the character in each of the scripts (Kanji or Kana), Yoshihara et al.'s (2017) results indicated that the masked priming effects are sensitive to the characteristics of the script a stimulus is presented in.

In the form preparation associative-cuing task, on the other hand, stimuli consisting of a mixture of Kana and Kanji words produced a mora-based preparation effect (e.g., かつら /**ka**.-tu.ra/ "wig," 歌舞伎 /**ka**-bu-ki/ "kabuki," 鞄 /**ka.ba.N**/ "bag") and no phoneme-based effect (e.g., かつら /**ka**.tu.ra/, くじら /**ku**.zi.ra/ "whale," 古墳 /**ko**-hu.N/ "ancient tomb"; Kureta et al., 2006). This data pattern was consistent with those observed in masked priming studies using Kana and Romaji stimuli (Verdonschot et al., 2011), although not with those using Kanji stimuli (Yoshihara et al., 2017).

In the form preparation associative-cuing task, however, Romaji stimuli *did* produce significant phoneme-based facilitation (Kureta et al., 2015); responses were faster when response words shared their initial letter/phoneme (e.g., *m-aguma*, *menko*, *moppu*) relative to the control condition (e.g., *maguma*, *robii*, *netsui*). This result would imply that form preparation effects are also susceptible to the script types of the stimuli. Kureta et al. (2015) concluded, however, that the phoneme-based effect observed for their Romaji stimuli did not indicate that the Japanese phonological unit is the phoneme. Instead, they suggested that the effect is likely to reflect a *strategy* employed by their participants. That is, the orthographic characteristics of Romaji were suspected to be the source of the strategy because the letters made it salient that words could be segmented at the letter/phoneme level. Indeed, the authors found in a postexperimental interview that "almost all of the participants were more or less aware that onset-phonemes were shared" (p. 56) in the critical condition. Participants could then utilize the orthographic information provided by the stimuli to enhance their task performance. For example, they could use the overlapping letter/phoneme as a cue for memorization of critical items (e.g., Alario et al., 2007). In addition, they might also prepare in advance the onset phonemes with the help of orthographic information (i.e., Romaji letters), using attentional resources (e.g., O'Seághdha & Frazer, 2014).

As such, although it is generally agreed that the phonological unit of the Japanese language is the mora, not all studies find results that directly support that conclusion. The lack of support for the conclusion in some studies might stem from the fact that there are multiple scripts in the Japanese language, and the effects of scripts can interact with the task demands in these two major tasks (e.g., Kureta et al., 2015; Yoshihara et al., 2017).

Recently, Verdonschot and Kinoshita (2018) proposed that the Stroop task would be an additional viable methodological tool for investigating the nature of the phonological unit. Participants in the most typical version of that task are asked to name the ink color in which a color word is printed (e.g., Stroop, 1935). Normally, color-naming responses are slower when the color word and its ink color are incongruent (e.g., the color word *green* printed in red ink) than when they are congruent (e.g., the color word *red* printed in red ink). Although this classic Stroop effect has been interpreted as reflecting control of verbal actions, Roelofs (2003) argued that Stroop effects are more suitably explained by a speech production model (WEAVER++; Levelt et al., 1999). Therefore, Verdonschot and Kinoshita based their investigation of the phonological unit on Roelofs's ideas and employed a variant of the Stroop task, the phonological Stroop task.

Previous research using the Stroop task had shown that color naming is faster when a (non-) word distractor and the ink color share their initial phoneme (e.g., "rez" in red ink) than when they do not (e.g., "rez" in blue ink; Coltheart, Woollams, Kinoshita, & Perry, 1999; Mousikou, Rastle, Besner, & Coltheart, 2015; Parris et al., 2019). Verdonschot and Kinoshita (2018) initially used this type of manipulation (i.e., the phonological Stroop task manipulation) to demonstrate that this task can be a valid tool for investigating the phonological unit. That is, they used Kana stimuli and tested whether this task would produce the same pattern of effects commonly observed in the form preparation and masked priming paradigms, namely, a significant mora-based effect but no phoneme-based effect (e.g., Kureta et al., 2006; Verdonschot et al., 2011). The results showed that such was indeed the case. That is, color-naming responses were significantly faster when the ink color and the Kana (nonword) distractor matched in their initial morae than when they did not (e.g., responses were faster for "ピャ" /**pi**.ya/ than for "リャ" /ri.ya/ when the ink color was pink /**pi**.N.ku/) whereas responses were *not* faster when the ink color and the Kana distractor matched in the initial phoneme than when they did not ("パャ" /**pa**.ya/ colored in pink was not faster than "ラャ" /ra.ya/ colored in pink). Verdonschot and Kinoshita took that pattern of results as credible evidence that the phonological Stroop task is an effective tool for evaluating the nature of the phonological unit in the Japanese language.

Verdonschot and Kinoshita (2018) further argued that one reason that the phonological Stroop task is effective in evaluating the phonological unit is because this task is unaffected by the orthographic properties of the stimuli. In making this

proposal, the authors referred to Roelofs's (2006) claim that the influence of spelling on word production emerges only "when it is relevant for the word production task at hand" (p. 36). Roelofs based this claim on his examination, using Dutch stimuli, of the form-preparation effects in three tasks: word reading, picture naming, and associative-cuing tasks. Specifically, he examined whether the form-preparation effects can be observed when the response words are phonologically related, but orthographically unrelated (e.g., kompas, colbert, cadeau "compass, jacket, present"). The results showed that the orthographic inconsistency led to the absence of a form-preparation effect only in the word-reading task (failing to replicate Damian & Bowers's, 2003, pattern in the form-preparation associative-cuing task). That is, an orthographic influence was observed only when the task required participants to directly read visually presented word targets. Based on these ideas, Verdonschot and Kinoshita reasoned that orthography would not affect spoken word production in a phonological Stroop task as "the goal of the task is to simply name the color" and "the word distractor is a priori irrelevant to the task at hand" (p. 414).

Under the assumption that the phonological Stroop task is free from orthographic influences, Verdonschot and Kinoshita (2018) then examined the phonological unit in Japanese using single Kanji characters as distractors (as well as the Katakana-transcribed distractors of those Kanji characters). In that experiment, a significant mora-based effect was observed: the color-naming responses were faster when the distractor Kanji character and the ink color matched in their initial morae (e.g., 右 /mi.gi/ colored in green /mi.do.ri/) than when they did not (e.g., 右 /mi.gi/ colored in white /si.ro/). Critically, this mora-based effect occurred even when the initial mora did *not* correspond to the *whole sound* of the initial character in the Kanji distractors. This result sharply contrasted with the masked priming results reported by Yoshihara et al. (2017) using Kanji compound words. Recall that in Yoshihara et al.'s study, a mora-based effect was not observed when the initial mora did not match the full phonology of the first character in the Kanji stimuli (e.g., 確保 /ka.ku-ho/ did not facilitate 火力 /ka-rjo.ku/). The different results in the masked priming and phonological Stroop tasks can be easily explained if one accepts Verdonschot and Kinoshita's key assumption that performance in the phonological Stroop task directly taps into phonological units because it is free from orthographic influences, but masked priming word reading does not because it is affected by orthography.

The present research

The purpose of the present research was to evaluate Verdonschot and Kinoshita's (2018) conclusion that the phonological Stroop task is not influenced by the orthographic

properties of the stimuli. There are two reasons why we believe empirical scrutiny of this conclusion is warranted. The first is that although the authors' logic as for why the task should be free from orthographic influences sounds straightforward, it only has a minimum of empirical support. Roelofs's conclusion, cited by Verdonschot and Kinoshita (2018), that "spelling of a word constrains word production only when it is relevant for the task at hand" (p. 36) specifically referred to word production in *Dutch*. In addition, the main goal of Roelofs's study was to examine whether the orthographic influence on *the form-preparation effect*, originally reported by Damian and Bowers (2003), would be replicated. It is therefore reasonable to assume that Roelofs' conclusion mainly referred to issues within the form-preparation paradigm using Dutch. His conclusion may very well not extend to experimental situations that employ a different task (i.e., a phonological Stroop task) and a different language (i.e., Japanese).

Second, although Verdonschot and Kinoshita (2018) observed, using Kana stimuli, a significant mora-based effect and no phoneme-based effect in the phonological Stroop task, the pattern also observed in the two commonly used tasks (Kureta et al., 2006; Verdonschot et al., 2011), that finding, by itself, might indicate that the Stroop task *is* affected by the orthographic properties of the stimuli. Specifically, the lack of phoneme-based effect for Kana stimuli could have been reflective of an orthographic influence. That is, it is possible that the characteristics of Kana characters (i.e., each representing a mora) made it difficult to observe phoneme-level effects. If the phonological Stroop task is truly immune to orthographic influences, there should be no phoneme-based effect even for Romaji stimuli, in which most letters denote a phoneme.

Because it would be important to test and verify the assumption that the phonological Stroop task is free of orthographic influence before subsequent theoretical implications based on the data derived from this task are discussed, the present research was designed to do just that.[1] Specifically, we compared the patterns of phonological Stroop effects when the stimuli were presented in Kana versus Romaji. We chose this contrast because Romaji would seem to present the most likely script to produce a phoneme-based effect and that script has not been investigated in the phonological Stroop task. In masked priming naming experiments, on the other hand, stimuli presented in Kana and Romaji produced the same patterns of results: no significant phoneme-based effect and a significant mora-based effect (Verdonschot et al., 2011). In a form

---

[1] It would be worth noting that Sumiya and Healy (2004, 2008) reported a larger bilingual Stroop interference effect for Katakana distractors compared with the effect for Hiragana distractors, implying an (apparent) orthographic influence. It is, however, likely that their results had a phonological, rather than an orthographic, basis because their Katakana distractors (e.g., ブルー /bu.ru.R/ "blue") were phonologically more similar to the ink colors (e.g., blue) than their Hiragana distractors (e.g., あお /a.o/).

preparation task, in contrast, the two scripts produced different patterns of results regarding any phoneme-based effect: Kana (and Kanji) stimuli yielded no effect (Kureta et al., 2015; Kureta et al., 2006), whereas Romaji stimuli yielded a significant effect (Kureta et al., 2015). Recall, however, that the significant phoneme-based effect for the Romaji stimuli in Kureta et al. (2015) was attributed to the use of strategic processing by the participants.

In the present research, we conducted three experiments using the phonological Stroop task. In Experiment 1, we examined whether Romaji stimuli would yield a significant phoneme-based effect. In Experiment 2, the same Romaji stimuli were transcribed into Katakana, to test whether the pattern of results would be different from that in Experiment 1. In Experiment 3, we again attempted to examine whether the results would be different between Kana and Romaji, using a new set of stimuli. Specifically, we transcribed the Katakana stimuli used by Verdonschot and Kinoshita (2018, Experiment 1) into Romaji. As it has been demonstrated that these stimuli yield no phoneme-based effect when written in Katakana, the question in Experiment 3 was simply whether there would be a significant phoneme-based effect when those Katakana stimuli were presented in Romaji. If the phonological Stroop task is not influenced by the orthographic properties of the distractors, then Romaji and Katakana stimuli should produce parallel results. That is, color-naming responses would not be faster when the distractor and the ink color match in their initial phonemes (i.e., what we will refer to as congruent trials) compared with when they do not (i.e., what we will refer to as incongruent trials), and this pattern should hold regardless of the script types the distractors are presented in (Romaji vs. Kana).

Experiment 1

Method

Participants Thirty-two undergraduate and graduate students from Waseda University participated in this experiment (age: 21.1 years on average, SD = 2.8). They were paid 500 JP¥ (about 4 US$) in exchange for their participation. All were native speakers of Japanese with normal or corrected-to-normal vision. Care was taken not to include very highly proficient Japanese–English bilinguals, as previous studies have shown that such bilinguals would use phoneme-sized phonological units when the targets involve alphabetic letters (Nakayama et al., 2016; Verdonschot & Masuda, 2020). In the present experiments, a participant sample size was chosen that would allow us to have at least 1,600 observations in each of the critical conditions (i.e., congruent vs. incongruent trials; see Brysbaert & Stevens, 2018). The present experiments were approved by the Ethics Review Committee on

Research with Human Subjects at Waseda University (Protocol #2018-216). In all the experiments, participants provided written informed consent before the experiment started.

Stimuli Sixty Japanese words (normally written in either Kanji or Kana) were selected as the base words (e.g., 頭 /a.ta.ma/ "head," クシ /ku.si/ "comb"). Their mean word frequency was 62 per million (based on Amano & Kondo, 2003). These base words were then transcribed using Romaji letters (e.g., *kushi*) to serve as distractors. All the distractors contained an initial phoneme which corresponded to the initial phoneme of a color name in Japanese (e.g., /k/, /m/, /s/, /h/, /a/). Each distractor was presented on two types of trials: congruent and incongruent trials. On the congruent trials, the initial phoneme of each distractor was the same as that of the color it was printed in; for instance, *kushi* was printed in yellow (i.e., /ki.i.ro/). On the incongruent trials, the initial phoneme of each distractor was not the same as that of the color it was printed in; for instance, *kushi* was printed in green (i.e., /mi.do.ri/). The incongruent trials were created by re-pairing the color and distractors from the congruent trials. Each distractor was presented twice— once in a congruent trial and once in an incongruent trial. Thus, there were in total 120 trials.

Six colors were used in the present experiment: /ki.i.ro/ "yellow," /mi.do.ri/ "green," /mu.ra.sa.ki/ "purple," /si.ro/ "white," /ha.i.i.ro/ "gray," and /a.ka/ "red." For the color red, the initial phoneme of the color name in Japanese /a.ka/ also corresponds to a mora. Based on the results of Verdonschot and Kinoshita (2018), we expected to observe a significant effect at least for this color. This condition was included as a manipulation check as well to gauge if the mora-based effect would be larger than any (pure) phoneme effects in the other color conditions.

Apparatus and procedure Participants were tested individually in a quiet room. The experiment was programmed using the DMDX software package (Forster & Forster, 2003). The experimental trials consisted of two blocks, each containing 60 trials. Within each block, half of the trials were congruent trials and the other half were incongruent trials. The same distractor was not presented twice within the same block. The proportions of the distractors' ink colors were equal in the two blocks. The presentation order of the two blocks and trials within each block were randomized across participants. Participants were asked to name the ink color as quickly and as accurately as possible.

Each trial started with a 50-ms 400 Hz beep signal. After the signal, a fixation mark (i.e., "+") was presented in the center of the CRT monitor for 816 ms and was then followed by a blank screen for 400 ms. Immediately after the blank screen, a colored word (i.e., a distractor) was presented until the participant named the ink color or 2,000 ms has elapsed. The intertrial interval was 816 ms. The distractors were

presented in lowercase letters on a black background (12-point Courier New font). Color-naming latency was measured from the onset of the distractor to the onset of the vocal response. Prior to the experimental trials, participants received 16 practice trials (using distractors not used in the main experiment) to familiarize themselves with the task.

Results

Data from two participants were excluded from the analyses—one due to a mechanical problem and the other due to a failure to follow the instructions. As a result, data from 30 participants were analyzed. Responses were preprocessed and manually corrected for voice-key errors via visual inspection of the speech waveform using CheckVocal software (Protopapas, 2007). Response latencies faster than 250 ms or slower than 1,300 ms were regarded as outliers and excluded from the analysis (0.6% of the data), resulting in 3,580 data points. Error responses (2.4%) were also excluded from the latency analyses, resulting in 3,495 data points. The mean response latencies and error rates are presented in Table 1.

We analyzed the response latency data using linear mixed-effects (LME) models (e.g., Baayen, Davidson, & Bates, 2008) and the lme4 package (Bates, Mächler, Bolker, & Walker, 2015b) available in R (Version 3.5.0; R Development Core Team, 2018). Following Verdonschot and Kinoshita (2018), we first examined the shape of the latency distribution for the correct responses (containing no outliers). Because a log transformation approximated a normal distribution better than a reciprocal inverse transformation did, we decided to apply a log transformation to the raw RTs (hereafter, *logRT*) to meet the assumption of normality. We used the lmerTest package in R (Kuznetsova, Brockhoff, and Christensen, 2017), to calculate the *p*-values with the degrees of freedom based on Satterthwaite's approximation. In the

analyses, Congruence (congruent vs. incongruent) and Color (/ki.i.ro/, /mi.do.ri/, /mu.ra.sa.ki/, /si.ro/, /ha.i.i.ro/, and /a.ka/) were treated as fixed-effect factors. The Congruence factor was contrast-coded as +0.5/−0.5. Simple-coding was applied to the factor Color and referenced to red (/a.ka/) to examine whether the mora-based effect is significantly larger than any pure phoneme-based effect. As a result, there were five types of contrasts for the Color factor: (a) /ki.i.ro/ vs. /a.ka/ (Contrast 1), (b) /mi.do.ri/ vs. /a.ka/ (Contrast 2), (c) /mu.ra.sa.ki/ vs. /a.ka/ (Contrast 3), (d) /si.ro/ vs. /a.ka/ (Contrast 4), and (e) /ha.i.i.ro/ vs. /a.ka/ (Contrast 5).

We determined the most parsimonious model (e.g., Bates, Kliegl, Vasishth, & Baayen, 2015a; Matuschek, Kliegl, Vasishth, Baayen, & Bates, 2017) by progressively entering random effect factors into the model if the model fit was improved significantly by doing so, based on the chi-squared likelihood ratio test. In the response latency analysis, only the random slope of Color for participants improved the model fit significantly (*p* < .001). However, the model including this random slope failed to converge and the available optimizers did not return similar values. We thus report the results from the model with only the random intercepts: "logRT ~ Congruence × Color + (1 | Participant) + (1 | Item)." The error analysis was conducted using the same procedure except that we used a generalized linear mixed-effect model, assuming a binomial distribution. As no random slope factor improved the model fit significantly (all *p*s > .05), the final model formula was "ERROR ~ Congruence × Color + (1 | Participant) + (1 | Item)." Note that because this final model failed to converge in the error analysis, we reran the model with all the optimizers and confirmed that all the optimizers returned similar values.[2]

In the analysis of response latencies, the effect of Congruence was significant, *estimated coef.* = −0.023, *SE* = 0.006, *t* = 4.137, *p* < .001, reflecting the fact that the color-naming responses were faster on congruent trials. That is, the responses were faster when the initial phoneme of the color name matched the initial phoneme of a distractor. All contrasts of the factor Color were also significant (all *p*s < .05), which means that the stimuli printed in red (/a.ka/) were responded significantly faster than the stimuli in the other colors across congruent and incongruent trials. Importantly, the interactions between Congruence and each contrast of Color were not significant (all *p*s > .05). This pattern means that the sizes of the phoneme-based and mora-based effects were not significantly different. In the analyses of errors, no effects were significant (all *p*s > .05).

Table 1 Mean color-naming latencies in milliseconds (ms) and error rates in percentages (%) with a net effect in Experiment 1

Color	Congruent		Incongruent		Stroop effect	
	RT	ER	RT	ER	RT	ER
/ki.i.ro/ (yellow)	583	0.7	603	1.3	20	0.6
/mi.do.ri/ (green)	627	3.0	635	3.0	8	0.0
/mu.ra.sa.ki/ (purple)	662	4.1	669	2.4	7	−1.7
/si.ro/ (white)	606	2.0	617	4.0	11	2.0
/ha.i.i.ro/ (grey)	685	2.3	687	5.4	2	3.1
/a.ka/ (red)	553	0.3	577	0.0	24	−0.3
Grand mean	619	2.1	631	2.7	12	0.6

Note. RT and ER stand for mean response latencies and error rates, respectively

---

[2] When a model failed to converge, we followed the recommendations made by the authors who constructed the *lme4* package (Bates et al., 2015b). In addition, we confirmed that the results did not change when a random slope was included.

Discussion

In Experiment 1, a significant phoneme-based effect was observed when distractors were presented in Romaji. Color-naming responses were significantly faster when the distractor and the name of ink color had the same initial phoneme (e.g., *kushi* printed in /ki.i.ro/ "yellow") than when they had different phonemes (e.g., *kushi* printed in /mi.do.ri/ "green"). In addition, although the initial phoneme of /a.ka/ (red) also corresponds to a mora and the effect was numerically larger for this color than for the other colors (see Table 1), there was no significant difference in the sizes of these effects. These results are inconsistent with the assumption that the phonological Stroop task is unaffected by the orthographic properties of distractors. If the task were free from orthographic influences, a phonemic effect should not have been observed with Romaji distractors.

Note that in Experiment 1 the color-naming responses were significantly faster when the distractors were printed in red than when they were printed in other colors. This result is in line with the results from some of the previous studies (e.g., Regan, 1978; Schadler & Thissen, 1981), although, as pointed out by Schadler and Thissen (1981), the color difference is rarely mentioned in color-naming studies. We must admit that it is not clear why such a difference was observed. We do, however, offer the following speculations here. First, Schadler and Thissen suggested that differences in brightness are often the source of color effects, as brighter colors (e.g., red) are perceived faster than the other colors. Alternatively, it may be that warm colors (e.g., red) are processed more efficiently than cool colors (e.g., blue, green), as real-life objects tend to be warm colored and backgrounds tend to be cool colored (e.g., Gibson et al., 2017). As noted just above, however, these are just speculations and providing an explanation of the red advantage is beyond the scope of the present research.

To solidify the conclusions derived from Experiment 1 concerning orthographic influences in the phonological Stroop task, we conducted Experiment 2. In that experiment, the distractors presented in Romaji in Experiment 1 were transcribed into Katakana. If the Katakana distractors produce no phoneme-based effect, it would be a successful replication of Verdonschot and Kinoshita's (2018) original observation using Katakana distractors. At the same time, however, such a data pattern (with respect to the results of Experiment 1) would indicate that the phonological Stroop task is *not* free from orthographic influences.

Experiment 2

Method

Participants Thirty undergraduate and graduate students from Waseda University participated in this experiment (age: 20.3 years on average, $SD = 1.3$). They were paid 500 JP¥ in exchange for their participation. All were native speakers of Japanese with normal or corrected-to-normal vision. None had participated in Experiment 1.

Stimuli, apparatus, and procedure The stimuli, apparatus, and procedure in Experiment 2 were the same with those in Experiment 1, except that the distractors were transcribed into Katakana characters (e.g., クシ /ku.si/).

Results

Responses were preprocessed and manually corrected for voice-key errors via visual inspection of the speech waveform using CheckVocal software (Protopapas, 2007). Response latencies faster than 250 ms or slower than 1,300 ms were regarded as outliers and excluded from both analyses (0.9% of the data), resulting in 3,569 data points. Error responses (3.1%) were also excluded from the latency analysis, resulting in 3,458 data points. The mean response latencies and error rates are presented in Table 2. The response latencies and errors were analyzed in the same way as in Experiment 1.

In the analysis of response latencies, the model including random slopes failed to converge and the available optimizers did not return similar values. We thus report the results from the model with only the random intercepts, which managed to converge. The effect of Congruence was not significant ($t = 0.30$, $p = .762$), indicating that, overall, color-naming responses were not faster in the congruent than in the incongruent trials. All contrasts of the factor Color were significant (all $ps < .001$), reflecting the fact that the stimuli presented in the color /a.ka/ (red) were named significantly faster than the stimuli in the other colors across congruent and incongruent trials. The interaction between Congruence and each contrast of Color was not significant (all $ps > .1$).

Table 2 Mean color-naming latencies in milliseconds (ms) and error rates in percentages (%) with a net effect in Experiment 2

Color	Congruent		Incongruent		Stroop effect	
	RT	ER	RT	ER	RT	ER
/ki.i.ro/ (yellow)	605	2.0	600	0.7	−5	−1.3
/mi.do.ri/ (green)	638	3.4	644	3.7	6	−0.3
/mu.ra.sa.ki/ (purple)	677	2.4	675	4.7	−2	2.3
/si.ro/ (white)	614	4.4	608	1.4	−6	−3.0
/ha.i.i.ro/ (grey)	697	7.5	692	6.9	−5	−0.6
/a.ka/ (red)	565	0.0	578	0.7	12	0.7
Grand mean	632	3.3	632	3.0	0	−0.3

Note. RT and ER stand for mean response latencies and error rates, respectively

In the error analyses, again, as models often failed to converge and optimizers returned widely differing values, we report results from the simplest model (i.e., the random intercept-only model). The analysis showed that there were no significant effects or interactions (all $ps > .05$).

To further examine whether orthography (i.e., script type) affects the phonological Stroop effect, we conducted combined analyses for the data from Experiments 1 and 2. In these analyses, the fixed factors were Congruence (congruent vs. incongruent), Color (/ki.i.ro/, /mi.do.ri/, /mu.ra.sa.ki/, /si.ro/, /ha.i.i.ro/, and /a.ka/), and Script (Romaji vs. Kana). The Congruence and Script factors were contrast-coded as +0.5/ −0.5. Simple coding was applied to the factor Color and referenced to red (/a.ka/) as in the preceding analyses. As a result, there were five types of contrasts for the Color factor: (a) /ki.i.ro/ vs. /a.ka/ (Contrast 1), (b) /mi.do.ri/ vs. /a.ka/ (Contrast 2), (c) /mu.ra.sa.ki/ vs. /a.ka/ (Contrast 3), (d) /si.ro/ vs. /a.ka/ (Contrast 4), and (e) /ha.i.i.ro/ vs. /a.ka/ (Contrast 5).

In the analyses of both response latencies and errors, although only the random slope of Color for participants improved the model fit significantly (both $ps < .001$), this model failed to converge, and the available optimizers did not return similar values. We thus report the results from the model with only the random intercepts: "logRT(ERROR) ~ Congruence × Color × Script + (1 | Participant) + (1 | Item)."

In the analysis of response latencies, the effect of Congruence was significant, *estimated coef.* = 0.012, *SE* = 0.004, $t = 3.211$, $p = .001$. All contrasts for the factor Color were also significant (all $ps < .05$). The interaction between Congruence and Contrast 3 of Color was significant, *estimated coef.* = −0.031, *SE* = 0.015, $t = −2.057$, $p = .042$. The interaction between Congruence and Contrast 5 of Color was also significant, *estimated coef.* = −0.036, *SE* = 0.017, $t = −2.133$, $p = .037$. These interactions indicate that the effect for the color red (/a.ka/) was significantly larger than the effect for the colors purple (/mu.ra.sa.ki/) and gray (/ha.i.i.ro/). Most importantly, the interaction between Congruence and Script was significant, *estimated coef.* = 0.022, *SE* = 0.008, $t = 2.802$, $p = .005$, indicating that there was a significant phoneme-based effect when the distractors were presented in Romaji, but not when they were presented in Katakana. No other effect was significant (all $ps > .05$). In the analyses of errors, no effects or interactions were significant (all $ps > .05$).

Discussion

When the distractors used in Experiment 1 were transcribed into Katakana characters, the phoneme-based effect vanished. The lack of an effect for Katakana distractors is in line with Verdonschot and Kinoshita's (2018) results showing no phoneme-based effect for Katakana nonword distractors.[3] It is thus clear that phonemic overlap does not affect color-naming responses when distractors are presented in

Katakana. This finding, however, contrasts with the findings from Experiment 1 in which the phoneme-based effect was significant when distractors were written in Romaji. This contrast shows that the phonological Stroop effect *is* modulated by the orthographic properties of distractors (i.e., script type).

Before considering these results further, we felt that it was important to replicate the phoneme-based effect for Romaji distractors observed in Experiment 1. In Experiment 3, we selected the Katakana nonword distractors used by Verdonschot and Kinoshita (2018) and transcribed them into Romaji. As their Katakana stimuli did not show a phoneme-based effect (although they did show a mora-based effect), if a significant phoneme-based effect is observed using the Romaji version of their stimuli, it would strongly suggest that the phonological Stroop task *is* modulated by the orthographic properties of distractors.

Experiment 3

Method

Participants Thirty-five undergraduate and graduate students from Waseda University participated in this experiment (age: 20.9 years on average, *SD* = 4.2). They were paid 500 JP¥ in exchange for their participation. All were native speakers of Japanese with normal or corrected-to-normal vision. None had participated in Experiment 1 or 2.

Stimuli We used the same stimuli and design as those used in Experiment 1 of Verdonschot and Kinoshita (2018), except that we transcribed their Katakana nonword distractors (all involving two characters) into Romaji. Three colors were used: /pi.N.ku/ "pink," /gu.ri.R.N/ "green," and /bu.ru.R/ "blue." Although native Japanese speakers might not use these color names frequently as these are originally loanwords (specifically, they might prefer to use /mi.do.ri/ and /a.o/ to refer to green and blue, respectively, instead of /gu.ri.R.N/ and /bu.ru.R/), they can produce these color names without difficulty.

There were in total 192 distractors. They were divided into four conditions: C-congruent, C-incongruent, CV-congruent, and CV-incongruent. The C-congruent distractors were 48 bimoraic nonwords (e.g., paya /pa.ja/ , gisa /gi.sa/, bowa /bo.wa/) containing an initial phoneme which corresponded to the initial phoneme of the color names (e.g., /p/, /g/, /b/). The C-incongruent distractors were generated from the C-congruent distractors by

---

[3] One may argue that there should be a significant congruence effect for red, because the initial phoneme of this color also corresponds to a mora. Indeed, when we directly analyzed the effect for red, the 12-ms difference in the response latencies (see Table 2) reached significance (*estimated coef.* = 0.023, *SE* = 0.011, $t = 2.150$, $p = .032$).

replacing the initial phonemes /p/ with /r/, /g/ with /s/, and /b/ with /n/ (e.g., raya /ra.ja/, shisa /si.sa/, nowa /no.wa/). The CV-congruent were 48 bimoraic nonwords (e.g., piya /pi.ja/, gusa /gu.sa/, buwa /bu.wa/) containing an initial mora which corresponded to the initial mora of the color names (e.g., /pi/, /gu/, /bu/). The CV-incongruent distractors were generated from the CV-congruent distractors by replacing the initial morae /pi/ with /ri/, /gu/ with /su/, and /tu/ with /nu/ (e.g., riya /ri.ja/, tsusa /tu.sa/, nuwa /nu.wa/). Note that we replaced one CV-congruent distractor buni /bu.ni/ with bune /bu.ne/ because this item was duplicated in Verdonschot and Kinoshita (2018).

Apparatus and procedure The apparatus and procedure in Experiment 3 were the same as those used in Experiments 1 and 2, except for the following aspects. The experimental trials consisted of four blocks, each containing 48 trials. The presentation orders of the four blocks and trials within each block were randomized across participants. The distractors were presented in lowercase letters on a white background (16-point Courier New font). Prior to the experimental trials, participants received 12 practice trials (using distractors not used in the main experiment) to familiarize themselves with the task.

Results

The data from one participant were excluded from the analyses because of excessive errors (>25%) and hence the data from 34 participants were analyzed. Responses were preprocessed and manually corrected for voice-key errors via visual inspection of the speech waveforms using CheckVocal software (Protopapas, 2007). Response latencies faster than 250 ms or slower than 1300 ms were regarded as outliers and excluded from the analysis (0.4% of the data), leaving 6,505 data points. Error responses (2.2%) were also excluded from the latency analyses, leaving 6,361 data points. The mean response latencies and error rates are presented in Table 3. The response latencies and errors were analyzed in a similar way as in Experiments 1 and 2. In Experiment 3, however, Onset (C vs. CV) and Congruence (congruent vs. incongruent)

were treated as fixed effect factors and contrast-coded as +0.5/−0.5.

In the analysis of response latencies, the random slope of Congruence for participants improved the model fit significantly ($p < .05$). The final model we report was thus "logRT ~ Congruence × Onset + (1 + Congruence | Participant) + (1 | Item)." The effect of Congruence was significant, *estimated coef.* = 0.067, SE = 0.009, t = 7.761, $p < .001$, indicating that the color-naming responses were faster on the congruent than on the incongruent trials. The effect of Onset was not significant ($t = -0.987$, $p = .325$). The interaction between Congruence and Onset was not significant ($t = 0.812$, $p = .418$), implying that the effect was not significantly different between the phoneme (C) and mora (CV) overlap conditions.

In the analyses of errors, the random slopes of Congruence and Onset for participants improved the model fit significantly (both ps < .05). The final model we report was thus "ERROR ~ Congruence × Onset + (1 + Congruence + Onset | Participant) + (1 | Item)." The effect of Congruence was significant, *estimated coef.* = 1.165, SE = 0.361, z = 3.228, $p = .001$, indicating that there were fewer errors on the congruent than on the incongruent trials. The effect of Onset was not significant ($z = 1.038$, $p = .299$). The interaction between Congruence and Onset was marginal, *estimated coef.* = -0.952, SE = 0.508, z = -1.875, $p = .061$, implying that the effect on the error rates was slightly smaller in the phoneme (C) overlap condition than in the mora (CV) overlap condition.

Discussion

We replicated our significant phoneme-based effect for Romaji distractors (Experiment 1) using the stimuli that were originally written in Katakana characters and yielded no phoneme-based effect in Verdonschot and Kinoshita's (2018) experiment. Our ultimate conclusion, therefore, is that the results in the phonological Stroop tasks are *not* free from orthographic influences.

General discussion

In the present research, we evaluated the potential usefulness of the phonological Stroop task as a tool for investigating the

Table 3 Mean color-naming latencies in milliseconds (ms) and error rates in percentage (%) with a net effect in Experiment 3

Onset congruence with color names	Congruent		Incongruent		Stroop effect	
	RT	ER	RT	ER	RT	ER
C (phoneme)	495	1.0	528	2.4	33	1.4
CV (mora)	487	1.7	531	3.8	44	2.1

Note. RT and ER stand for mean response latencies and error rates, respectively

phonological encoding process in speech production. A major advantage of the phonological Stroop task, according to Verdonschot and Kinoshita (2018), is that the task is unaffected by the orthographic properties of the stimuli and thus purely reflects underlying phonological encoding processes. Therefore, this task was deemed to be quite useful for investigating such processes in Japanese, potentially allowing a resolution of inconsistent results observed in previous investigations. To empirically test their assumption, we examined whether the script of the stimuli (Romaji vs. Kana) would modify the pattern of results. Quite straightforwardly, if the phonological Stroop task is free from orthographic influences, there should be parallel results for Romaji and Kana distractors. That is, there should be no phoneme-based effect when Romaji stimuli are used as distractors, just like there was no such effect when Verdonschot and Kinoshita used Kana distractors.

Contrary to Verdonschot and Kinoshita's (2018) proposal, however, the present results clearly showed that the phonological Stroop task *is affected* by orthography. In Experiment 1 using Romaji distractors, there was a significant phoneme-based effect: The color-naming responses were faster when the initial phoneme of the color name and the distractor matched (e.g., naming /ki.i.ro/ with the distractor *kushi*) than when they did not (e.g., naming /mi.do.ri/ with the distractor *kushi*). When the same distractors were transcribed into Katakana characters, however, the phoneme-based effect vanished (Experiment 2). Finally, in Experiment 3, we successfully replicated the phoneme-based effect using a new set of Romaji distractors that were created by transcribing the Katakana-written stimuli used in Verdonschot and Kinoshita's Experiment 1 which had yielded no phoneme-based effect. These results indicate that the orthographic properties of distractors (i.e., script types) do have an impact on phonological Stroop performance.

What is now needed, of course, is an explanation of how the orthography of the distractors affects the phonological Stroop effect. First, we do not believe that the present results imply that the phonological unit size in Japanese word production is the phoneme. Previous studies have repeatedly shown that native Japanese speakers do not employ a phonemic phonological unit (e.g., Ida, Nakayama, & Lupker, 2015; Kureta et al., 2006; Verdonschot et al., 2011; Verdonschot, Tokimoto, & Miyaoka, 2019), unless they are highly proficient Japanese–English bilinguals and English words are used as stimuli (Nakayama et al., 2016). In addition, although Kureta et al. (2015) also reported a phoneme-based form preparation effect using the associative-cuing task, they were likely correct in suggesting that the effect reflected a strategy employed by their participants, rather than that the Japanese phonological unit is the phoneme. Essentially, it seems unlikely that Japanese speech production employs the phoneme as the functional unit in the phonological encoding processes.

With respect to our own data, we suggest that the phoneme-based effect for Romaji distractors was due to a *strategy* used by participants, paralleling Kureta et al.'s (2015) claims. Specifically, we propose that, in the phonological Stroop task, Romaji stimuli made it salient that the initial phoneme overlapped between the distractor and color name on half the trials (i.e., the congruent trials). The participants then strategically paid attention to the distractor's initial phoneme. This process would, in turn, facilitate the color-naming responses on congruent trials and/or inhibit the color-naming on incongruent trials.[4] Such a relationship would likely be more difficult to notice when the distractors are presented in Kana (a mora-based script) and therefore no phoneme effect would be likely to appear in that situation.

The involvement of strategies, as was pointed out by Kureta et al. (2015), is also supported by the fact that phoneme-based effects were not observed for native Japanese speakers in a masked priming naming task using Romaji stimuli (Verdonschot et al., 2011), a task that is thought to be, at most, minimally affected by strategic factors (e.g., Forster, 1998). As such, our findings add a slight nuance to the argument by Roelofs (2006), who suggested that the influence of spelling on word production would appear only "when it is relevant for the word production task at hand" (p. 36). Possibly because Japanese words written in Romaji are slightly unusual, participants may have (strategically) paid extra attention to the spelling of distractors, which in turn led to the observed phoneme-based effects.

In addition, recall that the evidence of mora-based priming (i.e., for the color red) was fairly minimal in Experiment 2. This result may suggest that strategies also play a role in producing mora-based effects. That is, it is possible that the mora-overlap was not an effective cue for the participants in Experiment 2, because only the color red (/a.ka/) matched at the mora-level. If so, the participants would have had little motivation to strategically pay attention to the distractor's initial mora and, hence, the mora-based effect for red, although significant when examined by itself, was rather small.

The present findings, therefore, further suggest that the results in Verdonschot and Kinoshita (2018) may have been partly due to the use of a task-based strategy. As noted in the Introduction, those authors observed significant mora-based effects not only for Kana stimuli but also for Kanji stimuli (e.g., naming the color "green" /mi.do.ri/ was faster for stimuli such as 右 /mi.gi/). As both Kana and Kanji scripts are quite

---

[4] It is difficult to determine whether the phonological Stroop effect observed in the present study was facilitation or interference, because we did not include control trials in which the initial phoneme of a distractor was different from those in all the color names. Although this fact does not change our conclusion that the phonological Stroop effect is not free from orthographic influences, it would be important in future research to investigate whether the phonological Stroop effect is facilitation and/or interference (e.g., Hasshim & Parris, 2014; Marmurek, Proctor, & Javor, 2006).

frequently used in the Japanese writing system, it might have been difficult for their participants not to notice that the Kanji and Kana distractors had the same initial mora as the color names on the congruent trials, causing them to put a special emphasis on mora units.

Of course, Verdonschot and Kinoshita's (2018) finding of a mora-based effect for Kanji stimuli may not have been due to the use of a task-specific strategy alone. Rather, it may reflect the fact that Japanese native speakers tend to naturally segment the phonology of a word into its morae (e.g., Otake, Hatano, Cutler, & Mehler, 1993; see also Spagnoletti, Morais, Alegria, & Dominicy, 1989). That is, native Japanese speakers likely will extract a mora from a word with little difficulty regardless of whether the word is written in Kana or Kanji, as a result, often showing mora-based effects. In future research, it will be crucial to investigate to what extent the phonological Stroop effect is indeed modulated by the orthography of the distractors, per se, versus by a task-based strategy that the participants may have employed.

Before concluding, note also that the orthographic influence on the phonological Stroop task would likely not be limited to Japanese. That is, in most languages, the size of the phonological overlap needed to observe the phonological Stroop effect would likely correspond to the phonology possessed by each letter/character, the size of which is, of course, constant because those languages only involve one script. That is, a phoneme-based effect was found in English (e.g., Coltheart et al., 1999), a language in which most letters represent a phoneme, whereas, a (atonal) syllable-based effect was reported in Chinese (e.g., Li, Lin, Wang, & Jiang, 2013), a language in which each character stands for a syllable. Therefore, an obvious question would be whether a phoneme-based effect for Chinese might be observed in the phonological Stroop task if distractors are written in pinyin (i.e., Roman alphabetic letters used to transliterate the sound of a Chinese word), as was demonstrated in the form preparation paradigm (Li, Wang, & Idsardi, 2015).

In conclusion, the present experiments indicate that the phonological Stroop task is not free from orthographic influences. Our observation, of course, does not imply that the phonological Stroop task is not a useful tool for examining the phonological unit in speech production. To the extent that the Stroop phenomenon could be captured by a speech production model (e.g., Roelofs, 2003), that task may have some potential for shedding light on our understanding of speech production. For instance, Kinoshita and Verdonschot (2020) recently proposed the picture variant of the phonological Stroop task as another useful tool for investigating phonological encoding. We would, however, like to point out that it is important to empirically verify the strengths and weaknesses of any (relatively) novel experimental task before the data derived from that task are regarded as providing a strong basis for drawing theoretically based conclusions.

References

Alario, F.-X., Perre, L., Castel, C., & Ziegler, J. C. (2007). The role of orthography in speech production revisited. *Cognition, 102*, 464–475. doi:https://doi.org/10.1016/j.cognition.2006.02.002

Amano, N. & Kondo, K. (2003). *NTT Detabesu Shirizu: Nihongo No Goi-tokusei Dai 2-ki CD-ROM-ban* [NTT database series: Lexical Properties of Japanese, Vol.2, CD-ROM Version.] Tokyo, Japan: Sanseido.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*, 390–412. doi:https://doi.org/10.1016/j.jml.2007.12.005

Bates, D., Kliegl, R., Vasishth, S., & Baayen, R. H. (2015a). *Parsimonious mixed models*. Available from arXiv:1506.04967 (stat.ME).

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015b). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*, 1–48. doi:https://doi.org/10.18637/jss.v067.i01

Bi, Y., Wei, T., Janssen, N., & Han, Z. (2009). The contribution of orthography to spoken word production: Evidence from Mandarin Chinese. *Psychonomic Bulletin & Review, 16*, 555–560. doi:https://doi.org/10.3758/PBR.16.3.555

Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of Cognition, 1*(1):9, 1–20. doi:https://doi.org/10.5334/joc.10

Chen, J. Y., O'Séaghdha, P. G., & Chen, T. M. (2016). The primacy of abstract syllables in Chinese word production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42*(5), 825. doi:https://doi.org/10.1037/a0039911

Coltheart, M., Woollams, A., Kinoshita, S., & Perry, C. (1999). A position-sensitive Stroop effect: Further evidence for a left-to-right component in print-to-speech conversion. *Psychonomic Bulletin & Review, 6*, 456–463. doi:https://doi.org/10.3758/BF03210835

Damian, M. F., & Bowers, J. S. (2003). Effects of orthography on speech production in a form-preparation paradigm. *Journal of Memory and Language, 49*, 119–132. doi:https://doi.org/10.1016/S0749-596X(03)00008-1

R Development Core Team. (2018). R: A language and environment for statistical computing [Computer software]. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Forster, K. I. (1998). The pros and cons of masked priming. *Journal of Psycholinguistic Research, 27*, 203–233. doi:https://doi.org/10.1023/A:1023202116609

Forster, K. I., & Davis, C. (1991). The density constraint on form-priming in the naming task: Interference effects from a masked prime. *Journal of Memory and Language, 30*, 1–25. doi:https://doi.org/10.1016/0749-596X(91)90008-8

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, and Computers, 35*, 116–124. doi:https://doi.org/10.3758/BF03195503

Gibson, E., Futrell, R., Jara-Ettinger, J., Mahowald, Bergen, L., Ratnasingam, S., … Conway, B. R. (2017). Color naming across languages reflects color use. *Proceedings of the National Academy of Sciences of the United States of America, 114*, 10785–10790. doi: https://doi.org/10.1073/pnas.1619666114

Han, J.-I., & Verdonschot, R. G. (2019). Spoken-word production in Korean: A non-word masked priming and phonological Stroop task investigation. *Quarterly Journal of Experimental Psychology, 72*(4), 901–912. doi:https://doi.org/10.1177/1747021818770989

Hasshim, N., & Parris, B. A. (2014). Two-to-one color-response mapping and the presence of semantic conflict in the Stroop task. *Frontiers in Psychology, 5*, 1157. doi:https://doi.org/10.3389/fpsyg.2014.01157

Ida, K., Nakayama, M., & Lupker, S. J. (2015). The functional phonological unit of Japanese-English bilinguals is language dependent: Evidence from masked onset and mora priming effects. *Japanese Psychological Research, 57*, 38–49. doi:https://doi.org/10.1111/jpr.12066

Kinoshita, S., & Verdonschot, R. G. (2020). Phonological encoding is free from orthographic influence: Evidence from a picture variant of the phonological Stroop task. *Psychological Research* doi:https://doi.org/10.1007/s00426-020-01315-2

Kureta, Y., Fushimi, T., Sakuma, N., & Tatsumi, I. F. (2015). Orthographic influences on the word-onset phoneme preparation effect in native Japanese speakers: Evidence from the word-form preparation paradigm. *Japanese Psychological Research, 57*, 50–60. doi:https://doi.org/10.1111/jpr.12067

Kureta, Y., Fushimi, T., & Tatsumi, I. F. (2006). The functional unit in phonological encoding: Evidence for moraic representation in native Japanese speakers. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*, 1102-–1119. doi:https://doi.org/10.1037/0278-7393.32.5.1102

Kuznetsova, A., Brockhoff, P. B., Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*, 1–26. doi:https://doi.org/10.18637/jss.v082.i13

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Science, 22*, 1–75.

Li, C., Lin, C. Y., Wang, M., & Jiang, N. (2013). The activation of segmental and tonal information in visual word recognition. *Psychonomic Bulletin & Review, 20*, 773–779. doi:https://doi.org/10.3758/s13423-013-0395-2

Li, C., & Wang, M. (2017). The influence of orthographic experience on the development of phonological preparation in spoken word production. *Memory & Cognition, 45*, 956–973. doi:https://doi.org/10.3758/s13421-017-0712-5

Li, C., Wang, M., & Idsardi, W. (2015). The effect of orthographic form-cuing on the phonological preparation unit in spoken word production. *Memory & Cognition, 43*, 463–578. doi:https://doi.org/10.3758/s13421-014-0484-0

Marmurek, H. H., Proctor, C., & Javor, A. (2006). Stroop-like serial position effects in color naming of words and nonwords. *Experimental Psychology, 53*, 105. doi:https://doi.org/10.1027/1618-3169.53.2.105

Matuschek, H., Kliegl, R., Vasishth, R., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language, 94*, 305–315. doi:https://doi.org/10.1016/j.jml.2017.01.001

Meyer, A. S. (1990). The time course of phonological encoding in language production: The encoding of successive syllables of a word. *Journal of Memory and Language, 29*, 524–545. doi:https://doi.org/10.1016/0749-596X(90)90050-A

Meyer, A. S. (1991). The time course of phonological encoding in language production: Phonological encoding inside a syllable. *Journal of Memory and Language, 30*, 69–89. doi:https://doi.org/10.1016/0749-596X(91)90011-8

Mousikou, P., Rastle, K., Besner, D., & Coltheart, M. (2015). The locus of serial processing in reading aloud: Orthography-to-phonology computation or speech planning? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*, 1076–1099. doi:https://doi.org/10.1037/xlm0000090

Nakayama, M., Kinoshita, S., & Verdonschot, R. G. (2016). The emergence of a phoneme-sized unit in L2 speech production: Evidence from Japanese–English bilinguals. *Frontiers in Psychology, 23*, e46595, doi:https://doi.org/10.3389/fpsyg.2016.00175

O'Séaghdha, P. G. (2015). Across the great divide: Proximate units at the lexical-phonological interface. *Japanese Psychological Research, 57*, 4–21. doi:https://doi.org/10.1111/jpr.12074

O'Séaghdha, P. G., Chen, J.-Y., & Chen, T.-M. (2010). Proximate units in word production: Phonological encoding begins with syllables in Mandarin Chinese but with segments in English. *Cognition, 115*, 282–302. doi:https://doi.org/10.1016/j.cognition.2010.01.001

O'Séaghdha, P. G., & Frazer, A. K. (2014). The exception does not rule: Attention constraints form preparation in word production. *Journal of Experimental Psychology: Learning, Memory and Cognition, 40*, 797–810. doi:https://doi.org/10.1037/a0035576

Otake, T., Hatano, G., Cutler, A., & Mehler, J. (1993). Mora or syllable? Speech segmentation in Japanese. *Journal of Memory and Language, 32*, 258–278.

Parris, B. A., Sharma, D., Weekes, B. S. H., Momenian, M., Augustinova, M., & Ferrand, L. (2019). Response modality and the Stroop task: Are there phonological Stroop effects with manual responses? *Experimental Psychology, 66*, 361–367. doi:https://doi.org/10.1027/1618-3169/a000459

Protopapas, A. (2007). Check vocal: A program to facilitate checking the accuracy and response time of vocal responses from DMDX. *Behavior Research Methods, 39*, 859–862. doi:https://doi.org/10.3758/BF03192979

Regan, J. (1978). Involuntary automatic processing in color-naming tasks. *Perception & Psychophysics, 24*, 130–136. doi:https://doi.org/10.3758/BF03199539

Roelofs, A. (2003). Goal-referenced selection of verbal action: Modeling attentional control in the Stroop task. *Psychological Review, 110*, 88–125. doi:https://doi.org/10.1037/0033-295X.110.1.88

Roelofs, A. (2006). The influence of spelling on phonological encoding in word reading, object naming, and word generation. *Psychonomic Bulletin & Review, 13*, 33–37. doi:https://doi.org/10.3758/BF03193809

Roelofs, A. (2015). Modeling of phonological encoding in spoken word production: From Germanic languages to Mandarin Chinese and Japanese. *Japanese Psychological Research, 57*, 22–37. doi: https://doi.org/10.1111/jpr.12050

Schadler, M., & Thissen, D. M. (1981). The development of automatic word recognition and reading skill. *Memory & Cognition, 9*, 132–141. doi:https://doi.org/10.3758/BF03202327

Spagnoletti, C., Morais, J, Alegria, J., & Dominicy, M. (1989). Metaphonological abilities of Japanese children. *Reading and Writing, 2*, 221–244. doi:https://doi.org/10.1007/BF00377644

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*, 643–662. doi:https://doi.org/10.1037/h0054651

Sumiya, H., & Healy, A. F. (2004). Phonology in the bilingual Stroop effect. *Memory & Cognition, 32*, 752–758. doi:https://doi.org/10.3758/BF03195865

Sumiya, H., & Healy, A. F. (2008). The Stroop effect in English–Japanese bilinguals: The effect of phonological similarity. *Experimental Psychology, 55*, 93–101. doi:https://doi.org/10.1027/1618-3169.55.2.93

Verdonschot, R. G., & Kinoshita, S. (2018). Mora or more? The phonological unit of Japanese word production in the Stroop color naming task. *Memory & Cognition, 46,* 410–425. doi:https://doi.org/10.3758/s13421-017-0774-4

Verdonschot, R. G., Kiyama, S., Tamaoka, K., Kinoshita, S., La Heij, W., & Schiller, N. O. (2011). The functional unit of Japanese word naming: Evidence from masked priming. *Journal of Experimental Psychology: Learning, Memory and Cognition, 37*, 1458–1473. doi: https://doi.org/10.1037/a0024491

Verdonschot, R. G., Tokimoto, S., & Miyaoka, Y. (2019). The fundamental phonological unit of Japanese word production: An EEG study using the picture-word interference paradigm. *Journal of Neurolinguistics, 51*, 184–193. doi:https://doi.org/10.1016/j.jneuroling.2019.02.004

Warner, N., & Arai, T. (2001). The role of the mora in the timing of spontaneous Japanese speech. *The Journal of the Acoustical Society of America, 109*, 1144–1156. doi:https://doi.org/10.1121/1.1344156

Wong, A. W. K., Chiu, H. C., Wang, J., Wong, S. S., & Chen, H. C. (2019). Electrophysiological evidence for the time course of syllabic and sub-syllabic encoding in Cantonese spoken word production. *Language, Cognition and Neuroscience, 34*(6), 677–688. doi: https://doi.org/10.1080/23273798.2018.1562559

Yoshihara, M., Nakayama, M., Verdonschot, R. G., & Hino, Y. (2017). The phonological unit of Japanese kanji compounds: A masked priming investigation. *Journal of Experimental Psychology: Human Perception and Performance, 43*, 1303–1328. doi:https://doi.org/10.1037/xhp0000374

You, W.-P., Zhang, Q.-F., Verdonschot, R. G., (2012). Masked syllable priming effects in word and picture naming in Chinese. *PLOS ONE, 7*, e46595. doi:https://doi.org/10.1371/journal.pone.0046595