Psychological Research (2021) 85:1340–1347
https://doi.org/10.1007/s00426-020-01315-2

ORIGINAL ARTICLE

Phonological encoding is free from orthographic influence: evidence from a picture variant of the phonological Stroop task

Sachiko Kinoshita[1] · **Rinus G. Verdonschot**[2]

Received: 20 January 2020 / Accepted: 7 March 2020 / Published online: 20 March 2020

Abstract

The phonological Stroop task, in which the participant names the color of written distractors, is being used increasingly to study the phonological encoding process in speech production. A brief review of experimental paradigms used to study the phonological encoding process indicated that currently it is not known whether the onset overlap benefit (faster color naming when the distractor shares the onset segment with the color name) in a phonological Stroop task is due to phonology or orthography. The present paper investigated this question using a picture variant of the phonological Stroop task. Participants named a small set of line drawings of animals (e.g., camel) with a pseudoword distractor printed on it. Picture naming was facilitated when the distractor shared the onset segment with the picture name regardless of orthographic overlap (CUST–camel = KUST–camel < NUST–camel). We conclude that the picture variant of the phonological Stroop task is a useful tool to study the phonological encoding process, free of orthographic influence.

Introduction

An important process in speech production is phonological encoding (e.g., Levelt, Roelofs, & Meyer, 1999), that is, generating a phonetic plan from an abstract phonological form retrieved from lexical memory to drive articulation. More specifically, during *phonological encoding*, segmental information is placed in a metrical frame (that specifies, e.g., the number of syllables and stress pattern) to produce a prosodified "phonological word". This process is necessary, because the prosodic context of utterances varies across situations (e.g., question, expressing disbelief): without phonological encoding, the spoken output would sound flat, like synthesized speech. Phonological encoding is involved in any task that requires speech production—in everyday conversation, translating from one language to another, picture naming, or reading aloud.

There is still much about this process that is not yet known—for example, it is only beginning to be appreciated that the unit involved in phonological encoding (i.e., the "proximate unit") is not universal, but varies between languages. Whereas it is well established that the phoneme (segment) is the proximate unit in English and Dutch and other European languages, in Mandarin Chinese it is the atonal syllable (e.g., O'Seaghdha, Chen, & Chen, 2010) and in Japanese the mora (e.g., Kureta, Fushimi, & Tatsumi, 2006; Verdonschot & Kinoshita, 2018). With this need for more research in mind, we begin with a brief review of experimental paradigms that have been used to study the phonological encoding process. The review identifies the phonological Stroop task (to be described below) as one such task, but the review also identifies an as yet unresolved question with this task—namely, whether the origin of an effect that is taken as the evidence for a key characteristics of the phonological encoding process is phonological, or instead, orthographic. We then present an experiment to investigate this question.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00426-020-01315-2) contains supplementary material, which is available to authorized users.

✉ Sachiko Kinoshita
sachiko.kinoshita@mq.edu.au

Rinus G. Verdonschot
rinusverdonschot@gmail.com

[1] Department of Psychology, Macquarie University Centre for Reading (MQCR), Macquarie University, Sydney, NSW 2109, Australia

[2] Department of Oral and Maxillofacial Radiology, Institute of Biomedical and Health Sciences, Hiroshima University, 1-2-3 Kasumi, Minami-ku, Hiroshima 734-8553, Japan

Experimental investigation of the phonological encoding process

Meyer (1990, 1991) developed the first experimental paradigm to investigate the phonological encoding process, called the "implicit priming" (or form preparation) task. In this paradigm, participants first learn a small set of semantically related word pairs (e.g., fruit—pear), called a prompt and response, respectively. After learning these word pairs, the participant is asked to produce the response word (e.g., /pɛər/) when a prompt word (e.g., fruit) is presented. The critical manipulation is whether or not all response words within a block overlap regarding certain phonological features, for example, the initial phoneme. Blocks in which all response words share these characteristics (e.g., /pɛər/, /pɒnd/, /pɔn/) are termed homogeneous, while blocks in which the response words do not share these characteristics (e.g., /pɛər/, /reɪs/, /taʊn/) are termed heterogeneous. Using Dutch words, Meyer (1990, 1991) reported two important findings: (1) segment (phoneme) overlap at the beginning of the response words yields a response time benefit (i.e., homogeneous < heterogeneous), whereas overlap in non-initial position does not, and (2) the benefit increases with increasing overlap. These findings formed the key evidence for the serial nature (left-to-right incrementality) of the phonological encoding process, as well as the size of proximate unit—segments—in the Dutch language.

Since Meyer's (1990, 1991) pioneering work, the phonological encoding process has been studied with two other tasks, namely, the masked onset priming read aloud task and the phonological Stroop color naming task. In the former, a word (or a pronounceable nonword) target is presented to be read aloud. The target is preceded by a word (or nonword) prime presented briefly, usually no more than 50 ms, which is forward masked by a string of # signs and backward masked by the target (e.g., ####—save—SINK). Forster and Davis (1991) were the first to report that when the target is to be read aloud (but not when read silently as in the lexical decision task), a prime that shared *just the initial letter* with the target (e.g., save–SINK) facilitated the naming of the target relative to an unrelated, control prime (e.g., gave–SINK), and dubbed the effect the masked onset priming effect (MOPE). Subsequently, Kinoshita (2000), using English words and pseudowords, and Schiller (2004), using Dutch words, showed that the benefit due to a letter match was present at the beginning of the word and absent for the end overlap, just like the implicit priming task.

Another task that revealed the serial nature of phonological encoding process is the phonological Stroop task. When naming the color in which a color-neutral word (or a nonword) is written, response is faster when the word shares a phoneme with the color name to be produced (e.g., RAP written in red) than when it does not (e.g., FIT written in red). Coltheart, Woollams, Kinoshita, and Perry (1999) showed that this segment overlap effect is larger for the beginning of the word than the end of the word (e.g., KID written in red), and called it a position-sensitive Stroop effect.

Although the proponents of the dual-route cascaded model of reading (e.g., Coltheart et al., 1999, see also Mousikou, Rastle, Besner, & Coltheart, 2015) put forward both the MOPE and the position-sensitive Stroop effect as evidence for a serially operating grapheme–phoneme mapping process specific to reading, many others (e.g., Roelofs, 2004; Kinoshita, 2000; Schiller, 2004; Dimitropoulou, Duñabeitia, & Carreiras, 2010) have argued that it is more parsimonious to interpret the serial nature of the effects as originating in the phonological encoding process in speech production. A key argument against the grapheme–phoneme mapping account is that these serial effects are found not just with word (or pseudoword) stimuli written in the alphabetic script which plausibly involve the grapheme–phoneme mapping process for it to be named. For example, MOPE is found in a *letter naming* task with a letter prime where the name (e.g., "em" for M) cannot be generated via the grapheme–phoneme mapping process (Bowers, Vigliocco, & Haan, 1998, Experiment 2). Similarly, the position-sensitive Stroop effect is found with Japanese words written in *logographic kanji script* whose name cannot be generated via the grapheme–phoneme mapping process: Verdonschot and Kinoshita (2018) found that when naming the color of a word written in kanji, color naming was facilitated when the onset mora overlapped (e.g., naming /mu.ra.sa.ki/—purple in Japanese—was faster in the context of 娘 /mu.su.me/, meaning daughter, than in the context of 嵐 /a.ra.shi/, meaning storm). MOPE is ubiquitous, and is found when the to-be-named target is a picture (e.g., Schiller, 2008). Analogously, the "homogenous advantage" in the form preparation task is of equal magnitude whether the response items are pictures or written words (Roelofs, 2004). These findings limit the utility of the grapheme–phoneme mapping account as an explanation of onset overlap effects, and we will not discuss it further here.

In sum, this brief review indicated that three tasks—the implicit priming (form preparation) task, masked onset priming read aloud task, and the phonological Stroop task have all shown that performance is facilitated by the overlap in the onset segment, but not by the end segment. This "onset segment overlap benefit" forms the key evidence for both the serial nature (left-to-right incrementality) of, and the size of the unit used in the phonological encoding process.

Is the onset overlap benefit due to orthography?

An important question concerning the onset overlap benefit is whether the effect is orthographic, or phonological in origin. Due to the close correspondence between letters of the alphabet (e.g., T, B) and phonemes (/t/, /b/), orthography and phonology are often confounded in the alphabetic writing system, as words that share the onset phoneme also tend to share the same initial letter.

With the form preparation paradigm, using English words, Damian and Bowers (2003) reported that when onset segments (e.g., /k/) shared by the response words were spelled differently (e.g., camel, kidney), the advantage in response latency relative to the heterogeneous set was eliminated. This finding, however, turned out to be difficult to replicate: using Dutch words and French words respectively, Roelofs (2006, Experiment 3) and Alario, Perre, Castel, and Ziegler (2007) found no influence of orthography, at least when the response words were not presented as written words. These authors concluded that the effect of spelling reported by Damian and Bowers (2003) might not bear directly on the speech production processes, but rather on the memorization processes recruited by the prompt-response learning procedure. Along a similar line, Qu and Damian (2019) criticized this task as being of "questionable ecological validity" (p. 328) as a tool to study the online effects of orthography on speech production.

As for both the masked priming read aloud task and the phonological Stroop task, because they involve written primes/distractors, potentially, the origin of the onset overlap effects in these tasks could well be orthographic. However, the evidence from the masked priming read aloud task in fact shows that MOPE is completely phonological. Using Dutch words, Schiller (2007) reported that when the prime and target had orthographic and phonological onset overlap (e.g., consul–CAMPUS pronounced /kɔnsʏl/– /kɑmpʏs/), target word naming was facilitated relative to the unrelated control prime (e.g., houweel, "pickaxe" pronounced /hɑuˈwel/), and the size of the benefit did not differ for a prime that had only phonological overlap (e.g., koffie, "coffee", pronounced /kɔfi/). Schiller (2007) also observed no naming benefit with a prime that had only orthographic (and no phonological) overlap (e.g., cider–CAMPUS, pronounced /sidər/–/kɑmpʏs/).

To our knowledge, whether (alphabetic) orthography makes a contribution to the onset overlap effect in the phonological Stroop task has not been investigated. This is probably because suitable stimuli are difficult to find: in the Stroop task, the to-be-named stimuli are color names (e.g., red, blue) and most of them do not contain onsets that can be written with a different letter. For this reason, here we used pictures instead of colors as the to-be-named target, as pictures afford a greater range of stimuli to be used as the to-be-named target. The similarity of the Stroop color naming task and the PWI task has long been recognized, and Starreveld and La Heij (2017) presented a theoretical analysis of the two tasks, concluding that "picture–word interference is a Stroop effect" (the title of their paper).[1] In the present experiment, the similarity of the tasks was increased further by using a small set of to-be-named pictures from a single semantic category as in a Stroop color naming task. Using a small set of to-be-named pictures has been shown to speed up naming, and to make the latencies and the processes underlying the two tasks comparable (Shitova, Roelofs, Schriefers, Bastiaansen, & Schoffelen, 2016); this also made an effect that was present in the PWI task to disappear, just like the Stroop task (Geng, Schnur & Janssen, 2014). We therefore consider the present PWI task a picture variant of the Stroop task.

To summarize, the present experiment used a picture variant of the phonological Stroop task to investigate whether the onset overlap benefit is due to orthography, or phonology. Participants named a picture of animal (e.g., camel) with a pseudoword distractor printed on it, and we manipulated the overlap in the onset segment between the distractor and the picture name. It was either orthographically and phonologically congruent (e.g., CUST–camel), phonologically, but not orthographically congruent (e.g., KUST–camel), or neither orthographically nor phonologically congruent (e.g., NUST–camel). Relative to the last distractor type, based on previous findings of the "position-sensitive phonological Stroop effect" (Coltheart et al., 1999), we expected the orthographically and phonologically congruent distractor (e.g., CUST) to facilitate picture naming. The critical question was whether this facilitation is found for a distractor that shares only the phonology, and not the orthography (e.g., KUST).

---

[1] Starreveld and La Heij's (2017) paper was in direct opposition to Dell'Aqua, Job, Peressotti and Pascali (2007) who titled their paper "The picture–word interference effect is not a Stroop effect". In brief, Starreveld and La Heij noted that the results observed by Dell'Aqua et al. and taken as evidence for the dissociation between the two tasks have not been replicated in two later studies, and that the difference is likely to have been due to the methodological differences between the two tasks as they are standardly used. In particular, in the classic Stroop task, but not in the PWI task, only few targets selected from a single semantic category (colors) are used, and the distractors are also drawn from this category. Readers are referred to Starreveld and La Heij (2017) for further detail.

Experiment

Method

Participants

Twenty-five students (7 male, mean age 20.7, SD 4.15 years) from Macquarie University participated in the experiment in return for course credit. The experiment was approved by the Macquarie University Human Research Ethics Committee.

Design

The experiment used the picture–word interference (PWI) task, in which participants named a black-and-white line drawing of an animal with a written nonword distractor superimposed on it. The overlap in orthography and phonology between the onset of the picture name and the distractor was manipulated, and had three levels: (1) orthographically and phonologically congruent (O+P+), e.g., CUST–camel, (2) phonologically, but not orthographically, congruent (O−P+), e.g., KUST–camel, and (3) neither phonologically nor orthographically congruent (Control), e.g., NUST–camel. The dependent variables were response latency and error rate.

Materials

The to-be-named pictures were six black-and-white line drawings of animals (camel, giraffe, seal, rabbit, monkey and turtle) selected from Snodgrass and Vanderwalt (1980). The pictures were chosen on the basis that they were easily recognizable and easy to name (based on the normative data provided by Snodgrass and Vanderwalt (1980), name agreement ranged between 88 and 100%, with a mean of 94.6% for these stimuli), and the name started with a single consonant. Camel, giraffe and seal were chosen on the basis that the onset segment of their name can be written with a different letter (C/K for camel, G/J for giraffe, S/C for seal); they are referred to as the critical targets. Other pictures (rabbit, monkey and turtle)—whose names did not begin with a homophonic letter—are referred to as the filler targets. The fillers were included for two reasons: (1) to make the to-be-uttered response less predictable by increasing the response set size (from 3 to 6); and (2) to see if the onset overlap benefit found with the critical set of items (which can be written with homophonic letter) was of the same magnitude as the effect observed with the typically used stimuli.

The distractors were all pronounceable nonwords, four-letter long, monosyllabic, and started with a single consonant letter. There were three types of distractors: (1) O+P+, which shared the initial letter with the picture name (e.g., CUST–camel, SIMP–seal), (2) O−P+, distractors which were generated by replacing the initial letter of the O+P+ distractor with another letter that was homophonic (e.g., KUST from CUST, CIMP from SIMP), and (3) control, distractors which were generated by replacing the initial letter with another letter that was not homophonic (e.g., NUST from CUST, BIMP from SIMP). In all cases, the vowel was different from the vowel in the target (except for GISK/JISK/MISK–giraffe, due to an experimenter oversight), and we sought to use a variety of vowels. The distractor body (e.g., UST) had no phonological or orthographic overlap with the target name.[2] The critical target pictures were paired with three types of distractors (O+P+, O−P+ and control). The filler target pictures were paired with the O+P+ and control distractors (there were no O−P+ distractors for these pictures). For each picture, 10 distractors of each type were generated, thus in total there were 150 distractors (i.e., for the critical target pictures, there were 30 O+P+, 30 O−P+ and 30 control distractors; for the filler target pictures, there were 30 O+P+ and 30 control distractors). In addition, there were 12 distractors generated similarly, used for practice. The complete list of stimuli is presented in the Appendix available in Supplementary Materials.

Apparatus and procedure

Participants were tested individually, seated approximately 60 cm in front of a computer monitor, upon which the stimuli were presented. Each participant completed 150 test trials, in two blocks with a self-paced break between the two blocks. Each block contained the same number of picture targets and a representative number of three distractor types. A different pseudorandom order of trials was generated for the two blocks with the constraint that no picture target occurred in immediate succession. A practice block of 12 trials with the same picture targets as the test items but paired with different distractors preceded each test block.

Participants were instructed at the outset of the experiment that the task was a picture naming task. All six picture targets were shown, and the experimenter spoke the names (the written names were not shown). Participants were also told that each picture will have a written nonword superimposed on it, and their task was to ignore it and to name the picture as fast and accurately as possible.

Stimulus presentation and data collection were achieved through the use of the DMDX display system developed by

---

[2] Note that the orthographic (graphemic) overlap is not the same as letter overlap. Specifically, the vowel segment in “seal” is pronounced /iː/ and orthographically represented by the grapheme “ea”, and not “e” (/ɛ/) as in SELP. We deemed SELP, SELM, and SELG (for seal) as acceptable for this reason.

Table 1 Mean picture naming latencies (RT, in ms) and percent error rate (%*E*)

Distractor condition	Example	RT (%*E*)	Effect (vs. control)
Critical stimuli (target picture = camel)			
O+P+	CUST	594 (4.9%)	34 (1.3%)
O−P+	KUST	594 (5.6%)	34 (.6%)
Control	NUST	628 (6.2%)	
Filler stimuli (target picture = turtle)			
O+P+	TASH	583 (3.7%)	30 (2.1%)
Control	VASH	613 (5.8%)	

Forster and Forster (2003). Stimulus display was synchronized to the screen refresh rate (10.1 ms).

Each trial started with the presentation of a fixation signal ("+") for 250 ms, followed by a blank for 50 ms. A target picture was then presented as a black-and-white line drawing, with a nonword distractor written in red (in Arial 12 point font) superimposed on it. Targets were presented for a maximum of 2000 ms, or until the participant's response. The experimenter sat next to the participant, and monitored their oral response and noted down errors on a record sheet. Participants were given no feedback during the experiment.

Results

Correct picture naming latencies and error rates were analyzed using linear mixed effects (LME) modeling with subjects and items as crossed random effect factors, using the packages lme4 (version 1.1-17, Bates, Maechler, Bolker, & Walker, 2018), and lmerTest (version 3.0-1, Kuznetsova, Brockhoff, & Christensen, 2018) implemented in R (version 3.5.1, 2018–07–02, R Core Team, 2018). The fixed effect factor was distractor type. In the analysis of RTs, error trials and voice key trigger failures (28 trials, 0.7%) were excluded, and the RTs were log-transformed to meet the distributional assumptions of LME. We initially tested linear mixed-effects models with subject random slope on the fixed effect factor and subject and item random intercepts, and simplified the random effects structure if the model did not converge or the model fit was not improved by model complexity.

RTs

The mean correct RTs and error rates are shown in Table 1.

The critical targets were analyzed separately from the filler targets to test for the effects of orthographic and phonological overlap. The final statistical model we report here included as the fixed factor distractor type referenced to the O−P+ condition: logRT ~ distractor type + (1 | stimulus) + (1 | subj). Phonological overlap facilitated naming, as indicated by significantly slower response to the control condition relative to the O−P+ condition, $t = 2.456$, $B = 0.042$, $SE = 0.018$, $p < 0.01$. Critically, the O+P+ condition did not differ significantly from the O−P+ condition, $t = -0.521$, $B = -0.009$, $SE = 0.018$, $p = 0.603$, indicating that the orthographic overlap did not make any additional contribution.

To quantify the relative amount of evidence for the null difference between the O+P+ and O−P+ conditions, Bayes factor was calculated using the BayesFactor package (version 0.9.12-4.1, Morey & Rouder, 2018) comparing the model with the distractor type as the fixed effect factor as the denominator and the model without this factor (with subjects and stimuli as random effect factors) as the numerator. A Bayes factor is an odds ratio, with 1 indicating equal evidence for the two alternative hypotheses, and generally odds of 3 or greater are considered to provide "some evidence", greater than 10 to be "strong evidence", and odds greater than 30 to be "very strong evidence" (Jeffreys, 1961). The Bayes factor was eight in favor of the null hypothesis, indicating a moderate strength of evidence for the absence of an additional orthographic contribution.

In addition, we combined the critical targets and filler targets in the O+P+ and control conditions, and tested whether the difference between the O+P+ vs. control conditions differed for the critical and filler targets (picture names with onsets that can or cannot be written with an alternative homophonic letter). The distractor type factor (O+P+ vs. control) and the target status (critical vs. filler) were contrast coded (− 0.5, 0.5). The model we report is logRT ~ distractor type × target status + (1 | word) + (1 | subj). In this model, there was a significant effect of distractor type, $t = -5.050$, $B = -0.055$, $SE = 0.011$, $p < 0.001$. The Bayes factor for the distractor type effect was 325, indicating very strong evidence for the beneficial effect of phonological (and orthographic) onset match. Target status was non-significant, $t = -0.615$, $B = -0.007$, $SE = 0.01096$, $p = 0.54$, as was the target status by distractor type interaction, $t = -0.058$, $B = -0.001$, $SE = 0.022$, $p = 0.95$. The null interaction indicates that whether the onset segment can be written in multiple ways (as in the critical targets) or not (as in the filler targets)—i.e., the consistency of phonology-to-orthography mapping—had no impact on the size of the onset effect.

Error rate

Error rates were analyzed with generalized linear mixed effects model with subjects and stimuli as crossed random factors, using the logit function appropriate for categorical variables (Jaeger, 2008). As for RT, the model tested was: error rate ~ distractor type + (1 | subj) + (1 | stimulus), separately for the critical and filler targets. The error rates were generally low and differed little between conditions: For the critical targets, the model did not converge, and for the filler targets, the distractor type effect was non-significant, $Z = -1.608$, $p = 0.108$.

Discussion

The present experiment investigated the role of orthography in the onset segment match effect in a picture variant of the phonological Stroop task. The results were very clear. Relative to the unrelated control (e.g., NUST–camel), an onset-matching distractor (e.g., CUST or KUST) produced a sizable facilitation (~ 30 ms) in picture naming latency. Importantly, this facilitation was completely phonological, that is, orthographic overlap made no additional contribution (CUST–camel = KUST–camel).[3] The absence of orthographic contribution is further underscored by the fact that the consistency of phonology-to-orthography mapping of the onset segment (i.e., /k/ in "camel" can be written with C or K, whereas /t/ in "turtle" can be written with only the letter T) had little impact on the size of the onset overlap benefit. We take these findings to suggest that the PWI task can be adapted to serve as a tool to investigate the phonological encoding process in the PWI speech production task, free from the influence of orthography.

In making this claim, we should note that there are classic studies that used the PWI task to examine the contribution of orthography and phonology and reported results that at first glance appear inconsistent with the present finding. Specifically, like the present study, Lupker (1982, Experiment 2) used picture targets and pseudoword distractors and reported that relative to an unrelated control condition (e.g., distractor VOOSE with a picture of plane), phonologically similar distractors (e.g., TAIN) facilitated picture naming (680 ms vs. 665 ms). In apparent contradiction to the present result, this facilitation was substantially greater if the distractor additionally shared the spelling with the picture name (e.g., NANE, 627 ms). A similar pattern was observed with word distractors, with orthographically and phonologically related distractors (e.g., CANE–plane) facilitating picture naming more than a phonologically but not orthographically related distractor (e.g., BRAIN). How can these results be reconciled with the absence of orthographic influence in the present study?

The answer to this puzzle is that the phonologically (and orthographically) related distractor in Lupker's (1982) study shared the *rime* with the picture name, i.e., the overlap was at the *end* of the word and involved a vowel segment. In contrast, in the present experiment, the phonologically related distractors shared the consonant onset segment with the picture target (e.g., CUST/KUST–camel). Recall that the phonological encoding process is left-to-right incremental, and does not benefit from an end overlap. Thus, the phonological (and the additional orthographic) facilitation observed in Lupker's (1982) study could not have originated in the phonological encoding process. Indeed, Lupker's (1982) own interpretation was that "the locus of these effects appears to be the *name-retrieval process* (emphasis added) with orthographic and phonetic information from the word aiding in the search for the picture's name" (p. 349). Name retrieval precedes the phonological encoding process. Name retrieval is likely to have played a more limited role in the present experiment, as there were fewer picture targets (6 in the present study vs. 9 in Experiment 2 of Lupker, 1982), all of which had high name agreement,[4] and the participants were explicitly told of the name of each picture at the outset of the experiment. The fact that the picture naming latencies were considerably faster in the present study than in Lupker's is consistent with this possibility.[5] Moreover, in English, the relationship between orthography and phonology is much less predictable for vowels than for consonants; thus, the retrieval of the vowel sound would have been facilitated when the distractor and the picture name shared orthography as well as phonology. This would have been quite likely in Lupker's (1982) study, because some of the phonologically (but not orthographically) related distractors contained an irregular spelling of the rime, e.g., TOMB (for broom), CHOIR (for fire).

The point of the above discussion is to point out that the orthographic overlap between the written distractor and the to-be-named target can have an impact on a speech production task, but the origin of the effect may not be in the phonological encoding process. If the researcher's interest is specifically in the phonological encoding process (the process of segment-to-frame association), then methodological implementations to the PWI task to minimize the contribution of name retrieval process are recommended.

In conclusion, the present picture naming study showed that pseudoword distractors that share the onset segment with the picture target produced a sizable facilitation in picture naming, and that there was no additional contribution

---

[3] This pattern of finding (CUST = KUST < NUST when naming "camel") has since been replicated (Kinoshita & Mills, 2020). That study further found no difference between the three distractor conditions (CUST = KUST = NUST) when the response was a manual key press response and did not involve a speech response, consistent with the claim that the effect of onset overlap benefit originates in the phonological encoding process.

[4] In contrast, picture targets used by Lupker (1982) were selected from a children's coloring book, and the names of some of the pictures (e.g., "fire") may have been more ambiguous, which may have contributed to the greater role of the name retrieval process.

[5] In this context, we remind the readers that others (e.g., Geng et al., 2014; Shitova et al., 2016) have also noted a substantial reduction in picture naming latency when a small set of pictures is used repeatedly.

of orthographic overlap. We take these results to put forward the picture variant of the phonological Stroop task as a useful tool for investigating the phonological encoding process, a speech production process, free of orthographic influence.

Funding This study was funded by the Australian Research Council Discovery project scheme (Grant number DP140101199).

Compliance with ethical standards

Open practice statement The data and output of analyses reported in this paper are available at the Open Science Framework, at the following URL: https://osf.io/c7uyt/.

Conflict of interest Both authors declare they have no conflict of interest.

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the Macquarie University Human Research Ethics Committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

Alario, X. F., Perre, L., Castel, C., & Ziegler, J. C. (2007). The role of orthography in speech production revisited. *Cognition, 102*, 464–545.

Bates, D. M., Maechler, M., Bolker, B., & Walker, S. (2018). *Lme4: Linear mixed-effects models using 'Eigen' and S4. Version 1.1-17.*

Bowers, J. S., Vigliocco, G., & Haan, R. (1998). Orthographic, phonological, and articulatory contributions to masked letter and word priming. *Journal of Experimental Psychology: Human Perception and Performance, 24*, 1705–1719.

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.

Damian, M., & Bowers, J. (2003). Effects of orthography on speech production in a form-preparation paradigm. *Journal of Memory and Language, 49*, 119–132.

Dell'Acqua, R., Job, R., Peressotti, F., & Pascali, A. (2007). The picture–word interference effect is not a Stroop effect. *Psychonomic Bulletin & Review, 14,* 717–722.

Dimitropoulou, M., Duñabeitia, J. A., & Carreiras, M. (2010). Influence of prime lexicality, frequency, and pronounceability on the masked onset priming effect. *Quarterly Journal of Experimental Psychology, 63*, 1813–1837.

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 & Language, 30*, 1–25.

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

Geng, J., Schnur, T. T., & Janssen, N. (2014). Relative speed of processing affects interference in Stroop and picture–word interference paradigms: Evidence from the distractor frequency effect. *Language, Cognition and Neuroscience, 29*, 1100–1114.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59*, 434–446.

Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford: Oxford University Press, Clarendon Press.

Kinoshita, S. (2000). The left-to-right nature of the masked onset priming effect in naming. *Psychonomic Bulletin & Review, 7*, 133–141.

Kinoshita, S., & Mills, L. (2020). Phonological encoding in the oral but not manual Stroop task: Evidence for the role of a speech production process. *Journal of Experimental Psychology: Learning, Memory and Cognition*. https://doi.org/10.1037/xlm0000828.

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.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2018). *lmerTest: Tests in linear mixed effects models. Version 3.0-1*. https://CRAN.R-project.org/package=lmerTest. Accessed 02 July 2018.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*, 1–38.

Lupker, S. J. (1982). The role of phonetic and orthographic similarity in picture–word interference. *Canadian Journal of Psychology, 36*, 349–376.

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.

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.

Morey, R. D., & Rouder, J. N. (2018). BayesFactor: Computation of Bayes factors for common designs. R package version 0.9.12-4.1. Retrieved from https://CRAN.R-project.org/package=BayesFactor. Accessed 02 July 2018.

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 & Cognition, 41*, 1076–1099.

O'Seaghdha, 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.

Qu, Q., & Damian, M. (2019). Orthographic effects in Mandarin spoken language production. *Memory & Cognition, 47*, 326–334.

R Core Team. (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/. Accessed 02 July 2018.

Roelofs, A. (2004). Seriality of phonological encoding in naming objects and reading their names. *Memory & Cognition, 32*, 212–222.

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

Schiller, N. O. (2004). The onset effect in word naming. *Journal of Memory & Language, 50*, 477–490.

Schiller, N. O. (2007). Phonology and orthography in reading aloud. *Psychonomic Bulletin & Review, 14*, 460–465.

Schiller, N. O. (2008). The masked onset priming effect in picture naming. *Cognition, 106*, 952–962.

Shitova, N., Roelofs, A., Schriefers, H., Bastiaansen, M., & Schoffelen, J.-M. (2016). Using brain potentials to functionally localise Stroop-like effects in colour and picture naming: Perceptual encoding versus word planning. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0161052.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory, 6*(2), 174–215.

Starreveld, P. A., & La Heij, W. (2017). Picture-word interference is a Stroop effect: A theoretical analysis and new empirical findings. *Psychonomic Bulletin & Review, 24*, 721–733.

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.