OPEN ACCESS

University of BRISTOL

Zhang, Q., Zhu, X., & Damian, M. (2018). Phonological activation of category coordinates in spoken word production: Evidence for cascaded processing in English but not in Mandarin. *Applied Psycholinguistics*. https://doi.org/10.1017/S0142716418000024

Peer reviewed version

Link to published version (if available):
10.1017/S0142716418000024

Link to publication record in Explore Bristol Research
PDF-document

## University of Bristol - Explore Bristol Research
### General rights

**Phonological activation of category coordinates in spoken word production:**

**Evidence for cascaded processing in English but not in Mandarin**

Xuebing Zhu[3], Markus F. Damian[4], Qingfang Zhang[1,2]

[1]*Department of Psychology, Renmin University of China, Beijing, China*

[2]*Institute of Psychology, Chinese Academy of Sciences, Beijing, China*

[3]*Institute of Linguistic Studies, Shanghai International Studies University, Shanghai, China*

[4]*School of Experimental Psychology, University of Bristol, UK*

Word Count: 8,293 (main text excl. References)

RUNNING HEAD: Multiple phonological activation

**Abstract**

A central issue in spoken word production concerns how activation is transmitted from semantic to phonological levels. Recent evidence from studies of speakers of Western languages supports a cascaded view, according to which under certain circumstances, lexical candidates other than the target can activate their corresponding phonological properties. In the current study we investigated possible differences between English and Mandarin speakers concerning the degree of cascadedness in the production system, based on the broader recent claim that properties of word form encoding might differ according to languages. With English speakers (Experiment 1), we found that when activation of targets and semantic competitors was boosted via a manipulation of semantic context, then concurrently presented "mediated" distractor words (which were phonologically related to a semantic competitor) generated interference. Intriguingly, however, no such mediated priming was found in a parallel experiment with Chinese materials and Mandarin speakers (Experiment 2). These results suggest potential fundamental differences across the target languages in how activation is transmitted during lexical access.

Spoken production involves the timely coordination of semantic, syntactic and phonological properties of the words to be produced. In a word production task such as picture naming, recognition of the object induces activation of conceptual codes. Many models of spoken production make the assumption that a word's semantic and syntactic properties are represented in a "lemma" representation, and that semantically related lemmas are co-activated based on conceptual input. Activation then proceeds to a word form layer which represents a word's morphological and phonological properties (Dell, 1986; Jescheniak & Levelt, 1994; Levelt, Roelofs, & Meyer, 1999; Roelofs, 1992, 1997).

A long-standing debate concerns the temporal coordination of, and interaction between, those information types, and especially how activation is transmitted from lexical nodes to phonological encoding. In the literature, researchers have proposed two opposite possibilities: serial and non-serial models. *Serial discrete* models (e.g., Levelt et al., 1999) argue that for a given target word, only a single selected lemma spreads its activation to the phonological level, and semantic processing must be completed before phonological processing commences. By contrast, non-serial models such as *cascaded* models (e.g., Humphreys, Riddoch, & Quinlan, 1988; Morsella & Miozzo, 2002) propose that multiple lexical-semantic candidates which are co-activated during retrieval of the target word transmit activation to the phonological level. An even stronger form of non-seriality is proposed by *interactive* models (e.g., Dell, 1986) in which it is additionally assumed that transmission of activation between semantic and phonological processes is bidirectional, i.e., word form encoding can influence lexical-semantic retrieval. Critically, in serial models,

phonological activation is strictly restricted to the target word. By contrast, in non-serial models (both cascaded and interactive), phonological processing begins on the basis of early partial information provided by semantic processes, and any activated concepts and lemma nodes could transmit activation to their corresponding phonological nodes (in interactive models, activation additionally flows backward through the system). The present study investigates the possibility of multiple phonological activation differing across languages (in our case, English and Mandarin Chinese).

Findings from speech error studies have been taken to argue against a serial, and for an interactive, viewpoint (e.g., Dell & Reich, 1981; Dell, 1986). Recent findings from error-free spoken word production also argues against a strictly serial, and for a cascaded, processing mode. Much of this evidence is based on demonstrations of "multiple phonological activation"; i.e., entries other than the target are activated at the phonological level. For instance, in the "picture-picture interference" task, two line drawings are superimposed in different colors, and speakers name a target picture based on color while attempting to ignore the distractor picture. Morsella and Miozzo (2002) first demonstrated that when target and context picture were phonologically related (e.g., bed-bell), naming latencies were faster than when they were unrelated. The observation of phonological activation of context pictures has been replicated in English (Meyer & Damian, 2007) as well as in Spanish (Navarrete & Costa, 2005) and Dutch (Roelofs, 2008). This contradicts a central tenet of serial models of spoken word production (according to which phonological encoding should be restricted to the target only) and is in line with a cascaded view.

At the same time, cascading of activation from the semantic to the phonological level

does appear to be restricted. In a variation of the task in which a single colored picture is

presented, and speakers name either the object or the color, phonological facilitation is

found when colors are named (and objects ignored) but not when objects are named (and

colors ignored; e.g., Dumay & Damian, 2011; Kuipers & La Heij, 2009). This asymmetry

suggests that cascading is limited to the "primary" dimension (i.e., the object name) but

non-target properties such as color (or size, as shown more recently by Roux, Bonin &

Kandel, 2014) do not seem to be processed in a cascaded fashion. Hence, the evidence

suggests a "limited cascadedness" view of spoken word production.

Independent of the debate as to whether or not cascadedness is restricted to

particular target dimensions (see above), cascaded processing is itself probably quite

subtle, such that cascading of non-target properties is not detectable under all

circumstances. This appears to be the case in the popular picture-word interference (PWI)

task, in which participants are instructed to name target pictures while ignoring distractor

words superimposed on the target. A semantic relationship between target picture and

distractor (dog-cat) slows naming relative to an unrelated word, whereas a phonological

relationship (cat-cap) speeds up latencies (Glaser & Düngelhoff, 1984; Schriefers, Meyer, &

Levelt, 1990; Starreveld & La Heij, 1995). The possibility of cascadedness of non-target

properties can be investigated by using "mediated" distractors, i.e., words which are only

indirectly related to the target, such as the target picture "dog" paired with a distractor "can"

which is phonologically related to the semantic target coordinate "cat". If such mediated

distractors were to influence target naming, this would imply that not just the target (in this

case, "dog"), but also co-activated items (such as "cat") underwent phonological processing. However, mediated distractors appear not to affect target naming latencies in PWI tasks (Damian, 1998; Jescheniak, Hahne, Hoffmann & Wagner, 2006; but note that in the latter study, mediated effects were found in children) nor do mediated effects emerge in electrophysiological measures (Jescheniak, Hahne & Schriefers, 2003). This suggests that under typical circumstances, cascaded processing of non-target properties might be too weak to exert a measurable effect. By contrast, mediated priming has been shown with pictures with near-synonymous names such as Schäfer-Hirte (both meaning 'shepherd'; Jescheniak & Schriefers, 1998): when speakers named these pictures with the dominant name, then distractors which were phonologically related to the non-dominant name caused interference (see also Peterson & Savoy, 1998, for related findings with regard to synonyms). In the case of objects with near-synonymous names, it appears that both alternatives are phonologically encoded (hence implying cascadedness), with the incorrect alternative being primed by the mediated distractor and hence causing interference.

Assuming that in spoken word production, cascading of activation from semantic to phonological levels is genuine yet subtle, one could surmise that mediated effects will emerge more clearly if the degree of semantic and/or phonological overlap within mediated target/distractor combinations is increased. As outlined above, this is arguably the case with target objects with near-synonymous names such as Schäfer-Hirte (Jescheniak & Schriefers, 1998) for which the two alternatives are semantically virtually identical. An alternative strategy might be to boost activation at the semantic level via presentation of two semantically related pictures. Oppermann, Jescheniak, Schriefers, and Görges (2010)

asked speakers to name, cued by color, one of two spatially separated pictures on a computer screen; additionally, a spoken distractor word was presented. Distractors which were phonologically related to the non-target picture slowed down target naming latencies, but <u>only</u> when the two pictures were semantically related but not when they were unrelated. Accordingly, it is assumed that "there must be a semantics-sensitive mechanism that modulates the information flow in the conceptual-lexical system and that gates the amount of phonological activation (see Roelofs, 2008)" (p. 366).

Yet a different approach is to boost the magnitude of <u>phonological</u> activation via presentation of multiple distractor words. Abdel Rahman and Melinger (2008) superimposed two distractor words on a single target pictures, one of which sharing initial segments and the other sharing final segments with a word semantically related to the target. They found significant interference relative to an unrelated condition, consistent with the idea that when sufficiently primed, non-target alternatives are phonologically activated. Most recently, Melinger and Abdel Rahman (2013) reported that activating a concept associatively related to a target picture via form related distractor word pairs interfered with naming. For example, naming the target picture "pyramid" was slowed by the presence of the written distractors "camera" and "bagel" (form-related to "camel", an associate of "pyramid"), compared to an unrelated distractor combination. Overall, the evidence at present supports a theoretical framework in which activation transmission from semantic to phonological levels is cascaded.

*Word production in Western and non-Western languages*

While - as reviewed above - an increasing body of evidence supports cascaded models, the extant research is largely based on Indo-European languages such as English, German, Spanish and Dutch, and little attention has been paid to the possibility of a different phonological architecture in languages with non-alphabetic scripts. For instance, in the WEAVER model (Roelofs, 1997) word form encoding includes parallel access to phonological segments as well as retrieval of suprasegmental information. The two types of information are subsequently sequentially merged and associated with syllables in an incremental fashion. However, this architecture might not be universal across languages, and it has recently been suggested (O'Seaghdha, 2015; O'Seaghdha, Chen & Chen, 2010) that languages differ in the "proximate unit" of phonological encoding (i.e., the primary selectable unit below the word level). In Western languages, phonological segments constitute proximate units, and many priming effects demonstrated in experimental tasks (such as phonological facilitation in PWI tasks, e.g., (Glaser & Düngelhoff, 1984; Schriefers et al., 1990) are based on segmental overlap. Intriguingly, however, similar segmental manipulations in experiments conducted on Chinese individuals have resulted in null findings and instead priming effects are observed with syllabic manipulations only (e.g., Chen, Chen, & Dell, 2002; O'Seaghdha et al. 2010; Chen, Lin, & Ferrand, 2003; You, Zhang, & Verdonschot, 2012; Verdonschot, Nakayama, Zhang, Tamaoka, & Schiler, 2013). This suggests that in non-Western languages such as Mandarin and Cantonese, syllables (rather than segments) constitute the proximate units. Roelofs (2015) recently provided a first attempt to computationally model such differences between languages concerning phonological encoding.

Potential differences between languages concerning the architecture of phonological

encoding might have consequences for the issue of how activation is transmitted from the

semantic to the phonological level. And indeed, we have recently reported work which

highlights such a difference. For instance, a possible strategy of tackling the serial vs.

non-serial issue is to apply additive-factors logic (Sternberg, 1969) in a PWI experiment. By

factorially crossing semantic and phonological overlap between targets and distractors, one

can test for statistical additivity between the two variables. Additivity would be accounted

for more easily with a serial model, whereas a statistical interaction would be more in line

with a non-serial notion of lexical access. A substantial number of previous studies have

demonstrated non-additivity (i.e., a statistical interaction between semantic and

phonological overlap) in such experiments, across various Western languages (e.g.,

English: Damian & Martin, 1999; Taylor & Burke, 2002; Dutch: Starreveld & La Heij, 1995;

French: Bonin & Fayol, 2000), and this pattern is generally taken to support a non-serial

notion. However, with Chinese materials and Mandarin speakers, we (Zhu, Damian &

Zhang, 2016) found a strictly additive relation between semantic and (syllable-based)

phonological relatedness. This pattern is more in line with a serial notion of lexical access,

and underscores a potential important difference between Western and non-Western

languages concerning phonological encoding.

Further evidence for this claim comes from studies based on electroencephalography

(EEG), a method which allows to track access to representational stages before an overt

response has commenced. Using a picture-word interference task combined with EEG,

Dell'Acqua, Sessa, Peressotti et al. (2010) found significant effects of semantic and

phonological relatedness in the time window of 250-450 ms post picture onset, with peak

latencies of semantically related distractors (320 ms) coincided temporally with those of

phonologically related distractors (321 ms). By contrast, with Chinese speakers, we (Zhu,

Damian & Zhang, 2015) found a semantic effect in a time window of 250-450 ms which was

followed by a phonological effect in a much later time window, at 450-600 ms. Hence EEG

results from Mandarin speakers suggest a temporal dissociation between semantic and

phonological stages in Chinese, i.e., a serial/sequential pattern. Overall, these results lend

some support to the possibility that phonological encoding might differ in important aspects

between Western and non-Western languages such as Mandarin.

*The present study*

In the present study, we further tackled the issue of seriality vs. cascadedness in

spoken word production, as well as potential differences between Western and

non-Western languages, with a novel approach. For our experiments, we reverted back to

the use of "mediated" distractors in PWI tasks (distractors which are phonologically related

to a semantic competitor of the target object), a manipulation which as summarized above

typically results in null findings (e.g., Damian, 1998) but under certain circumstances might

render interference, namely when semantic (e.g., Oppermann et al., 2010) or phonological

(e.g., Abdel Rahman & Melinger, 2008) activation is increased.

In our experiments we manipulated activation at the semantic level with a

manipulation of "semantic blocking". In this task, participants repeatedly name a small set

of objects within an experimental block. Item sets within blocks are chosen such that they

either belong to the same semantic category ("homogeneous" condition), or each picture

comes from a different category ("heterogeneous"). The typical finding is that naming

latencies are longer in the homogeneous than in the heterogeneous condition (Abdel

Rahman & Melinger, 2007; Aristei et al., 2011; Belke, Meyer & Damian, 2005; Damian & Als,

2005; Damian, Vigliocco, & Levelt., 2001). Pictures in the same semantic category cause

additional activation of the related concepts and their corresponding lexical items, which

enhances the competition and delays lemma selection. However, the exact mechanism by

which this effect occurs remains somewhat controversial. This issue is less important for

our present purposes because irrespective of what exactly causes the effect, it is safe to

assume that in homogeneous contexts, targets and their semantic competitors are more

highly activated than in heterogeneous contexts.

We combined the semantic blocking manipulation with the use of mediated (and

other) distractors (note that Aristei et al., 2011, also used semantic blocking in conjunction

with a PWI manipulation, but in their case, with semantically related distractor words). On

each trial, a target picture was paired with one of three kinds of visual distractor words:

unrelated, phonologically related, and mediated (i.e., phonologically related to a semantic

competitor). See Figure 1 for a sketch of a "homogeneous" block with various distractors

superimposed on target objects. In line with numerous previous findings in the PWI

literature, we expected a facilitation effect from phonologically related distractors, and this

effect should be of comparable magnitude in semantically heterogeneous and

homogeneous contexts. The critical manipulation concerned the "mediated" distractors

which on any given trial were unrelated to the target word, but phonologically related to a

semantic competitor. Based on existing findings (e.g., Damian, 1998) we predicted that in

semantically <u>heterogeneous</u> blocks, mediated distractors should not affect target naming latencies. By contrast, in semantically <u>homogeneous</u> blocks, targets as well as semantic competitors undergo heightened activation, and hence we predicted that here mediated distractors might slow down target naming latencies because distractors of this type will further prime an (already pre-activated) potential competitor. We tested these predictions in Experiment 1 with English speakers.

(Insert Figure 1 here)

As highlighted in the previous section, there is preliminary support for the claim that phonological encoding in non-Western languages such as Mandarin might differ from Western languages. Specifically, Zhu et al. (2016) presented evidence for a more strictly serial notion of lexical access in Mandarin speakers, and Zhu et al. (2015) demonstrated that the time course of phonological encoding, as measured by EEG, differed from previous results obtained with Western speakers. A possible prediction from these findings is that the mediated effects that we predicted in Experiment 1, restricted to semantically homogeneous contexts, should generally be absent with Mandarin speakers. We tested this possibility in Experiment 2.

**Experiment 1**

*Method*

*Participants.* Twenty-four students (5 male; average age 21.4 years; range 18-31 years) from the University of Bristol were paid or received course credit for their participation. All were native English speakers and had normal or corrected-to-normal vision.

*Materials and Design.* Sixteen black and white line pictures were selected from the standardized picture database of Snodgrass and Vanderwart (1980), including four objects in each of four semantic categories (body parts, furniture, tools, and vehicles). Objects were combined into sets of four in order to form four "homogeneous" and four "heterogeneous" sets: in "homogeneous" blocks, all four pictures were from the same semantic category, whereas in "heterogeneous" blocks, one picture came from each semantic category. All pictures had monosyllabic names.

Each target picture was paired with three distractor words. A phonologically related word was chosen that shared one or more word-initial phonemes (58% segmental overlap on average) with the picture name (i.e., target: "train"; distractor: "trail"). A mediated word was chosen which shared one or more word-initial phonemes (54% shared segments on average) with a picture name from the same semantic category as the target (i.e., target: "train", distractor: "bulb", which is phonologically related to "bus"). An unrelated distractor word was selected that stood in no obvious relationship to the target (i.e., "goose" as a distractor). Distractors in each condition were statistically matched for length in phonemes and letters, and written frequency based on the normative information reported in the Neighborhood Watch program (Davis, 2005). An additional four pictures were selected as practice items.

The experimental design included the variables Semantic Context (homogeneous and heterogeneous) and Distractor Type (phonologically related, mediated, and unrelated) as within-participants and within-items variables. Eight blocks which each contained four target pictures, presented repeatedly, were constructed. Within each block, each target

was repeated four times under each distractor type. Therefore, each of the four targets

occurred 12 times for a total 48 trials in each block. The order of items within one block was

pseudo-randomized for each participant with the constraint that a particular target and the

first phoneme of a target name was never the same on consecutive trials. Four

homogeneous blocks and four heterogeneous blocks were constructed, yielding a total of

384 trials. Homogeneous and heterogeneous blocks were presented in alternating orders

and the order of different block lists was counterbalanced according to a Latin square

design.

We assessed the degree of semantic relatedness between semantic competitors of

mediated distractors and target names, and the one between unrelated words and target

names by 16 native English speakers (8 males, age from19 to 26 years old) who did not

take part in the main experiment 1. Target picture names were paired with their

corresponding semantic competitors of mediator distractor words and unrelated distractor

words, respectively. The word pairs were presented in random order, and pictures from

same category were avoided in the consecutive trials. The word pairs were rated on a

5-point scale, with 5 indicating that word pairs were highly semantically related and 1

indicating that word pairs were semantically unrelated. The average degree of semantic

relatedness was 1.74 (SD = 0.42) with a range of 1.25 to 2.69 between unrelated

distractors and target names, and was 3.63 (SD = 0.45) with a range of 3.00 to 4.26

between semantic competitors and target names across subjects. Paired *t-test* indicated a

significant difference between two semantic relatedness degrees, $t(15) = 14.85$, $p < .001$.

*Apparatus.* Stimuli were presented via an IBM-compatible computer on a 17-in. monitor

using DMDX 3.0 (Forster & Forster, 2003). Pictures were standardized to a size of

approximately 6 x 6 cm and displayed at the center of the screen. Distractor words were

presented in 22-point Times New Roman font, centrally superimposed on the target pictures.

Naming latencies were measured from target onset using a digital voice-key.

*Procedure.* Participants were tested individually in a sound-proof room. They were

seated approximately 60 cm from a computer screen. Participants were asked to familiarize

themselves with the experimental stimuli by viewing each target for 3,000 ms with the correct

name printed underneath. Then, participants were instructed to name individual target

pictures as fast and accurately as possible while attempting to ignore superimposed

distractor words. In a subsequent practice block, four additional pictures paired with unrelated

distractor words were presented twice. Then, eight experimental blocks of 48 trials each were

carried out.

Each trial involved the following sequence: A fixation point (*) was presented in the

middle of the screen for 500 ms, followed by a blank screen for 500 ms. Then, the target

picture and distractor word were presented simultaneously on the screen. Target pictures and

distractor words disappeared when participants initiated a voice response. An inter-trial

interval of 1,500 ms was included in each trial. The experiment took about 40 minutes in total.

*Result*

Data from incorrect responses and other responses caused by microphone errors

(3.4%), naming latencies longer than 2,000 ms or shorter than 200 ms (0.3 %), and those

deviating by more than three standard deviations from a participant's mean (1.28%) were

removed from the RT analyses. Furthermore, it is well known from previous studies that the

effect of semantic context differs between first presentation of an object within a block and all

subsequent presentations, with either little or no effect on first presentation (e.g., Aristei et al.,

2011; Belke et al., 2005; Damian & Als, 2005), or a facilitatory effect (e.g., Abdel Rahman &

Melinger, 2007). Because this pattern is of little interest to the purpose of our current study,

we removed data from the first presentation of an object paired with each kind of distractor

within each block.[1] Error rates were low (overall 1.6%) and thus were not analyzed further.

Figure 2 presents mean picture naming latencies and standard errors by semantic

context and distractor type. As expected, a sizeable semantic blocking effect is visible, as

well as priming from phonologically related distractors. Mediated distractors have no effect in

the semantically heterogeneous context, but generate numerical interference in the

homogeneous context.

（Insert Figure 2 here）

Analyses of variance (ANOVAs) were conducted on the response latencies, with

participants ($F_1$) or items ($F_2$) as random factors and Semantic Context and Distractor Type

as within-participants and within-items variables. A significant effect of Semantic Context was

found, $F_1$ (1, 23) = 27.07, $MSE$ = 1,445, $p < .001$, $\eta_p^2$ = .54; $F_2$ (1, 15) = 20.62, $MSE$ = 1,330, $p$

$< .001$, $\eta_p^2$ = .58, and a significant effect of Distractor Type, $F_1$ (2, 46) = 104.9, $MSE$ = 479, $p$

$< .001$, $\eta_p^2$ = .82; $F_2$ (2, 30) = 56.96, $MSE$ = 597, $p < .001$, $\eta_p^2$ = .79. The interaction between

Semantic Context and Distractor Type was significant, $F_1$ (2, 46) = 3.86, $MSE$ = 519, $p < .05$,

$\eta_p^2$ = .14; $F_2$ (2, 30) = 9.61, $MSE$ = 136, $p < .001$, $\eta_p^2$ = .39.

In order to assess the effects of phonological and mediated overlap separately, as well

as to explore potential interactions with the semantic context manipulation, we conducted two

additional analyses in which the unrelated condition and one of the two related conditions

(phonologically or mediated) was included whereas the other related condition was removed.

First, ANOVAs with Semantic Context (homogeneous vs. heterogeneous) and Phonological

Relatedness (related vs. unrelated) as within-participants and within-items factors revealed a

significant effect of Semantic Context, $F_1$ (1, 23) = 13.43, $MSE$ = 1,164, $p$ < .001, $\eta_p^2$= .37; $F_2$

(1, 15) = 11.01, $MSE$ = 1,015, $p$ < .01, $\eta_p^2$= .42, a significant effect of Distractor Type, $F_1$ (1,

23) = 126.74, $MSE$ = 407, $p$ < .001, $\eta_p^2$= .85; $F_2$ (1, 15) = 57.19, $MSE$ = 616, $p$ < .001, $\eta_p^2$= .79,

but no interaction between Semantic Context and Phonological Relatedness, $F_1$ (1, 23) < 1,

$MSE$ = 207, $p$ > .1; $F_2$ (1, 15) < 1, $MSE$ = 181, $p$ > .1. Planned $t$-tests for the phonological

facilitation in the homogeneous context ($M_{\text{diff}}$ = -45 ms) were significant, $t_1$ (23) = -7.59, $p$

< .001; $t_2$ (15) = -5.37, $p$ < .001, and so were $t$-tests for phonological facilitation in the

heterogeneous context ($M_{\text{diff}}$ = -47 ms), $t_1$ (23) = -12.07, $p$ < .001; $t_2$ (15) = -9.5, $p$ < .001.

Hence, we found the expected phonological facilitation, and this effect was independent of

semantic context.

Second, ANOVAs with Semantic Context (homogeneous or heterogeneous) and

mediated relatedness (related vs. unrelated) as within-participants and within-items factors

revealed a significant effect of Semantic Context, $F_1$ (1, 23) = 28.03, $MSE$ = 1124, $p$ < .001,

$\eta_p^2$= .55; $F_2$ (1, 15) = 24.97, $MSE$ = 887, $p$ < .001, $\eta_p^2$= .63, as well a significant effect of

Distractor Type, $F_1$ (1, 23) = 16.06, $MSE$ =377, $p$ < .001, $\eta_p^2$= .41; $F_2$ (1, 15) = 8.14, $MSE$ =

491, $p$ < .05, $\eta_p^2$= .35. Importantly, a significant interaction of Semantic Context and Distractor

Type was obtained, $F_1$ (1, 23) = 6.44, $MSE$ = 503, $p$ < .05, $\eta_p^2$= .22; $F_2$ (1, 15) = 30.21, $MSE$ =

68, $p$ < .001, $\eta_p^2$= .67. Planned $t$-tests for the mediated effect in the homogeneous context

($M_{diff}$ = -27 ms) were significant, $t_1$ (23) = 3.82, $p$ < .00; $t_2$ (15) = 4.82, $p$ < .001, but $t$-tests for

the mediated effect in the heterogeneous context ($M_{diff}$ = 4 ms) were not, $t_1$ (23) = .92, $p$ = .37;

$t_2$ (15) = .72, $p$ = .48. Hence, mediated distractors had an effect only in a semantically

homogeneous context.

To evaluate the effect size of the phonological and the mediated effect, we calculated

their *Cohen's d* for the homogeneous and heterogeneous conditions separately. For the

phonological effect, *Cohen's d* was 0.76 (heterogeneous) and 0.60 (homogeneous). For the

mediated effect, *Cohen's d* was 0.06 (heterogeneous) and 0.31 (homogeneous).

*Discussion*

The main findings of Experiment 1 were as follows. First, in line with numerous recent

studies (e.g., Abdel Rahman & Melinger, 2007; Aristei et al., 2011; Belke et al., 2005; Damian

et al., 2001), a semantic context effect was observed: pictures were named more slowly in

semantic homogeneous than in heterogeneous blocks. This effect likely arises from boosted

semantic activation among response items when named in homogeneous, compared to

heterogeneous, context. Second, we found a facilitation effect from phonologically related

distractors, a finding which is again predicted by numerous existent studies (e.g., Damian &

Martin, 1999; Schriefers et al., 1990; Starreveld & La Heij, 1995) and shows that English

speakers benefit from segment-sized (i.e., sub-syllabic) phonological overlap between

distractor and target. Third and most importantly, we found an interfering effect of mediated

distractors, but this effect was present only in a semantically homogeneous, but not in a

heterogeneous context. This finding demonstrates multiple phonological activation (i.e.,

phonological activation of a non-target lexical entry) but <u>only</u> when targets and competitors

are already pre-activated via the semantic context in which they occur. Hence, the results add

to the evidence suggesting that activation transmission to the phonological level is restricted.

The combination of PWI with semantic blocking provides a task which is sensitive to

otherwise relatively weak multiple phonological activation.

Experiment 2 used the same experimental manipulation, but now with Chinese

materials and native Mandarin speakers. The semantic context effect, as well as the

facilitatory effect of phonologically related distractors, should be similar across languages.

However, based on previous results (e.g., Zhu et al., 2015, 2016) we predicted that the

mediated effect which emerged in the first experiment, but only under a semantically

homogeneous context, should be absent in the second experiment.

## Experiment 2

*Method*

*Participants.* Twenty undergraduate students (10 male; average age 21.4 years;

range 19-25 years) from Beijing Forest University and China Agricultural University were

paid for their participation. All were native Mandarin Chinese speakers and had normal or

corrected-to-normal vision.

*Materials and Design.* Twenty-five black and white line pictures were selected from a

standardized picture database in Chinese (Zhang & Yang, 2003), including five objects

from each of five semantic categories (animals, body parts, clothing, fruits, and tools). All

pictures had disyllabic names. The objects were combined into sets of five in order to form

five homogeneous and five heterogeneous blocks, with the latter ones including one item in

each semantic category.

Each target picture was paired with three disyllabic distractor words. A phonologically related word was chosen that shared the first syllable but not the tone with the picture name (i.e., target: 袋鼠, /dai4shu3/, *kangaroo;* distractor*:* 歹徒, /dai3tu2/, *gangster*). A mediated word was chosen which shared the first syllable but not the tone with the first character of a picture name from the same semantic category as the target (i.e., target "kangaroo", distractor: 席子, /xi2zi5/, "mat", which is phonologically related to 犀牛, /xi1niu2/, "rhinoceros"). An unrelated word was selected that stood in no obvious relationship to the target (i.e., 枕头 /zhen3tou2/, *pillow*). Distractors in each condition were statistically matched for number of strokes and written frequency based on normative information reported in the database of the Chinese Lexicon (2003). An additional two drawings were selected as practice items.

The experimental design included the variables semantic context (homogeneous and heterogeneous) and distractor type (phonologically related, mediated, and unrelated) as within-participants and within-items variables. Ten blocks which each contained five pictures were constructed. Within each block, each target was repeated four times in each distractor type. Therefore, each of the five targets occurred 12 times for a total 60 trials in each block. The order of items within one block was pseudo-randomized for each participant with the constraint that a particular target and the first phoneme of a target name was never the same on consecutive trials. Five homogeneous blocks and five heterogeneous blocks were constructed, yielding a total of 600 trials. Homogeneous and heterogeneous blocks were presented in alternating orders and the order of different block lists was counterbalanced according to a Latin square design.

We also assessed the degree of semantic relatedness between semantic competitors of mediated distractors and target names, and the one between unrelated words and target names by 16 native Chinese speakers (4 males, age from18 to 45 years old) who did not take part in the main experiment 2. Identical rating procedure as the one in English was used. The average degree of semantic relatedness was 1.61 (SD = 0.42) with a range of 1.00 to 2.64 between unrelated distractors and target names, and was 3.75 (SD = 0.52) with a range of 3.08 to 4.68 between semantic competitors and target names across subjects. Paired t-test indicated a significant difference between two semantic relatedness degrees, $t(15) = 15.25$, $p < .001$. For semantic competitors of the mediator distractor words, independent *t-test* indicated there was no significant difference between English and Chinese speakers, $t(30) = -0.73$, $p = 0.47$, reflecting that the degree of semantic relatedness between target names and semantic competitors were similar in English and Chinese.

*Apparatus.* The experiment was performed using E-Prime Professional Software (Version 1.1; Psychology Software Tools). Pictures were standardized to a size of approximately 6 x 6 cm and displayed at the center of the screen. Distractor words were presented in 30-point Song font, centrally superimposed on the target pictures. Naming latencies were measured from target onset using a digital voice-key, connected with the computer via a PST Serial Response Box.

*Procedure.* The procedure was identical to Experiment 1, with the exception that 10 experimental blocks of 60 trials each were presented. The experiment took about 60 minutes in total per participant.

*Results*

Data from incorrect responses (0.9%), other responses such as mouth clicks (0.6%),

naming latencies longer than 2,000 ms or shorter than 200 ms (0.008 %), and those deviating

by more than three standard deviations from a participant's mean (1.51%) were removed

from all analyses. As in Experiment 1, data of the first presentation of an object paired with

three different kinds of distractors within each block were removed from the analysis. Error

rates were low (overall < 1%) and thus were not analyzed further.

Figure 3 presents mean picture naming latencies and standard errors by semantic

context and distractor type. As in the first experiment, the expected semantic blocking effect,

as well as facilitation from phonologically related distractors, is visible. By contrast, mediated

distractors appear to have little or no effect, and this is the case both in the semantically

homogeneous and heterogeneous context.

(Insert Figure 3 here)

Analyses of variance (ANOVAs) were conducted on the response latency means,

with participants ($F_1$) or items ($F_2$) as random factors and Semantic Context and Distractor

Type as within-participants and within-items variables. A significant effect of Semantic

Context was found, $F_1$ (1, 19) = 78.87, *MSE* = 924, *p* < .001, $\eta_p^2$= .81; $F_2$ (1, 24) = 70.40,

*MSE* = 1321, *p* < .001, $\eta_p^2$= .75, as well a significant effect of Distractor Type, $F_1$ (2, 38) =

29.62, *MSE* = 209, *p* < .001, $\eta_p^2$= .61; $F_2$ (2, 48) = 28.64, *MSE* = 276, *p* < .001, $\eta_p^2$= .54. No

interaction was found between these two variables, $F_1$ (2, 38) = .39, *p* = .68; $F_2$ (2, 48) = .57,

*p* = .57. Because the absence of this interaction contrasts with the results from the first

experiment, and in order to establish the likelihood of the null hypothesis being true, we

further conducted a Bayesian analysis with the method suggested by Rouder, Morey,

Speckman, and Province (2012) using JASP software (Love, Selker, Marsman et al., 2015).

The results revealed that the model with only the two main effects was superior to the full

model including the interaction, with a Bayes factor of $BF_{10} = 5.36$, which implies that the

null hypothesis is more than five times more likely than the alternative hypothesis.

According to the classification suggested by Jeffreys (1999), this constitutes "substantial"

evidence for the null hypothesis (i.e., the finding of no interaction is true).

   In order to assess the effects of phonological and mediated overlap separately, planned

t-tests analogous to those conducted in the first experiment were carried out. These showed

that the phonological facilitation effect in the semantically homogeneous context ($M_{diff}$ = -23

ms) was significant, $t_1 (19) = -5.18$, $p < .001$; $t_2 (24) = -5.40$, $p < .001$, as was the effect in the

semantically heterogeneous context ($M_{diff}$ = -21 ms), $t_1 (19) = -5.04$, $p < .001$; $t_2 (24) = -7.86$, $p$

$< .001$. By contrast, mediated distractors had no effect under the semantically homogeneous

context ($M_{diff}$ = 1 ms), $t_1 (19) = .42$, $p = .68$; $t_2 (24) = .41$, $p = .68$, nor under the semantically

heterogeneous context ($M_{diff}$ = -2 ms), $t_1 (19) = - .34$, $p = .74$; $t_2 (24) = - .20$, $p = .84$. To further

support the null finding concerning the effects of mediated distractors, we conducted

Bayesian analysis using JASP software as mentioned above. Using the

Jeffreys–Zellner–Siow (JZS) Bayes-factor paired $t$-test (Rouder, Speckman, Sun, Morey, &

Iverson, 2009), our result showed a Bayes factor $BF_{10}$ of 3.99 for the mediated effect in

homogeneous blocks, and 4.08 for heterogeneous blocks, suggesting that the null hypothesis

is approximately four times more likely than the alternative hypothesis. Again, the results

provide substantial support for the null hypothesis over the alternative.

To evaluate the effect size of the phonological and the mediated effect, we calculated

their *Cohen's d* for the homogeneous and heterogeneous conditions separately. For the

phonological effect, *Cohen's d* was 0.28 (heterogeneous) and 0.31 (homogeneous). For the

mediated effect, *Cohen's d* was 0.02 (heterogeneous) and 0.01 (homogeneous).

*Discussion*

The results of Experiment 1, with English speakers, had suggested multiple

phonological activation, as indicated by an interfering impact of mediated distractors which

was restricted to a semantically homogeneous context. The goal of Experiment 2 was to

determine whether this was also the case with Mandarin speakers. The results showed the

expected semantic context effect (we are not aware of previous studies which had used

semantic blocking in Chinese, but *prima facie* this effect should be independent of response

language) as well as phonological facilitation from related distractors (Wong & Chen, 2008;

Zhang, Chen, Weekes, & Yang, 2009; Zhang & Yang, 2005; Zhao, La Heij, & Schiller, 2012).

Importantly, however, no mediated effect was found, a finding which clearly diverges from the

results of the first experiment with English speakers. Although semantic activation was

boosted when objects were named in a context of other items belonging to the same category

(hence giving rise to the semantic blocking effect), mediated distractors which were

phonologically related to one of the semantic alternatives showed no effect. The absence of a

mediated effect in Mandarin cannot be attributed to insensitivity of the task we used since we

obtained mediated effects in Experiment 1 (subtle differences in design and materials

between the two experiments will be discussed in detail below). Hence, we argue that the

discrepancy between the results obtained from English and Mandarin speakers concerning

mediated effect arises from differences with regard to phonological encoding.

**General Discussion**

In the experiments reported here, we revisited the issue of information transmission from semantic to phonological levels in spoken word production. Combining the semantic blocking paradigm with a picture-word interference manipulation, the critical question was whether "mediated" distractors (words which are phonologically related to a semantic competitor of the target object) exert an effect on target naming latencies. If so, the results could lend further support to the claim that information transmission in spoken production is "cascaded" (e.g., Kuipers & La Heij, 2009; Morsella & Miozzo, 2002; Navarrete & Costa, 2005) and at least under some circumstances involves the activation of multiple phonological entries. In Experiment 1 we used English speakers and materials, and (besides the expected effects of semantic context and phonologically related distractors) we found mediated priming, but only in the semantically homogeneous, but not in the heterogeneous, context. In Experiment 2, we used Chinese materials and native Mandarin speakers in a design which was otherwise largely analogous to the first study. Again, we found semantic context effects and facilitation from phonologically related distractors. Critically, however, mediated priming was found under neither the semantically homogeneous, nor the heterogeneous, context.

The English results are generally in line with those from previous studies conducted on speakers of Western languages. As summarized in the Introduction, there is accumulating evidence for a "cascaded" view of lexical access in speaking: it is principally possible for non-target lexical entries to cascade activation to the phonological level, but

cascadedness is generally so subtle that it cannot be detected in all tasks and

circumstances. Hence, for instance, mediated distractors in PWI tasks (words which are

phonologically related to a semantic competitor of the target object) show no effect on

target naming under usual circumstances (e.g., Damian, 1998). Only when either semantic

or phonological activation is boosted relative to a "standard" case (e.g., when two

semantically related pictures are presented, as in Oppermann et al., 2010, or when a single

picture is paired with two phonologically related distractors, as in Abdel Rahman & Melinger,

2008) do mediated distractors show an effect. This is in line with our findings from English

speakers: mediated distractors had no effect in semantically heterogeneous blocks, but

they generated interference when presented in semantically homogeneous blocks in which

semantic activation of targets and competitors is presumably increased.

The results from Mandarin speakers are perhaps more surprising. In a largely

analogous experiment, the expected semantic blocking and phonological facilitation effects

were found. This is reassuring as there is no obvious reason why these effects should be

affected by target language. However, no mediated effects were found in this experiment.

At face value, the absence of mediated effects might be interpreted as evidence for serial

information transmission in Mandarin. And if so, our results would suggest a fundamental

difference between the target languages with regard to phonological encoding: weak

cascadedness in English, but strictly serial transmission in Mandarin.

We highlight the fact that the absence of a mediated effect in spoken Mandarin is in

fact fully compatible with two sets of results that we recently reported: Zhu et al. (2016)

factorially crossed semantic and phonological relatedness in a PWI task with Mandarin

speakers, and in contrast to numerous previous findings from speakers of Western

languages, the two types of relatedness exerted a strictly additive relationship in Mandarin.

Based on additive factors logic, this can be interpreted as indicating serial/discrete

information transmission between semantic and phonological levels. Zhu et al. (2015)

provided further evidence for a serial model via EEG and showed that with Mandarin

speakers, semantic and phonological stages emerged in sequential corresponding time

windows, which conflicts with comparable EEG studies conducted on speakers of Western

languages where both stages appeared largely at the same time. Both sets of results are in

line with the absence of a mediated effect in our present Experiment 2, and point toward a

serial transmission mode in Mandarin spoken word production.

Is it possible that the difference observed in the mediated condition between English

and Chinese arose at the conceptual level? Perhaps speakers of the two languages exhibit

differences in which they mentally represent the semantic categories which we used (tools,

furniture, body parts, etc.), and the presence or absence of a mediated effect hinges on the

underlying conceptual representations. However, studies on such semantic categories

have suggested strong similarity across languages (e.g., Bowerman, 1973; Brown, 1973;

Slobin, 1970, 1973), and accordingly, the chosen categories and exemplars in our

experiments successfully evoked the semantic blocking effect numerously reported before,

and in both languages. Indeed, the semantic effect in our task was numerically more

pronounced in Chinese than in English, which should render it more likely that mediated

effects should be observed in the former compared to the latter language. But instead, we

found mediated effects only in English but not in Chinese. This makes a conceptual origin

of the differences in the mediated condition unlikely.[2]

It is acknowledged that there are subtle differences in materials and manipulations

across the two studies which might hamper a clear interpretation of the results. These are

as follows. In Experiment 1 we used four semantic categories with four examples each; in

Experiment 2 there were five categories with five exemplars. This variation is unlikely to

make a difference and is within the limits of the existing literature (e.g., Damian et al., 2001,

used a 5 x 5 design whereas Damian and Als, 2005, used 4 x 4). Due to the difference in

categories/exemplars between the two experiments, but given the identical repetition of

targets within each block (12), this resulted in 384 trials per participant in the first

experiment, but 600 trials in the second. It is unlikely that the overall length of the

experiments would have influenced the results (e.g., Maess, Friederici, Damian, Meyer, &

Levelt, 2002, reported a semantic blocking experiment with a total of 1,200 trials).

Additionally, form overlap in our study was manipulated in terms of both orthography

and phonology for the English stimuli, but exclusively in terms of phonology for the Chinese

stimuli.[3] This discrepancy arose from constraints on stimulus selection; ideally, one would

use either English distractors which are only phonologically but not orthographically related,

or Chinese distractors which additionally share orthographic properties such as the first

character with the target. Unfortunately, both strategies are difficult to implement: in

alphabetic languages such as English, sound and spelling are necessarily confounded so it

is difficult or impossible to find word pairs which are phonologically but not orthographically

related. In Chinese, if a distractor and a target share the first orthographic character,

semantic associations between the two are unavoidable. One could also consider

repeating our two experiments with spoken, rather than written, distractors, but again this is

problematic as due to the prevalent homophony in Mandarin, isolated spoken words are

often difficult to disambiguate. Considering these potential factors, the different findings in

English and Chinese should be interpreted cautiously and need to be investigated further.

A further variation which is difficult to avoid is that in the English experiment, targets

and distractors were monosyllabic, whereas in Chinese they were disyllabic. This arises

from the statistics of the target languages: within the constraints of the semantic blocking

paradigm, it would be difficult to identify adequate disyllabic targets in English, or

monosyllabic targets in Chinese. Also, phonological overlap was segmental in the English

experiment, but syllabic in Chinese. Again, this inconsistency cannot be avoided: because

targets are monosyllabic in English, phonological overlap is necessarily sub-syllabic and

segmental. In Mandarin, by contrast, subsyllabic segmental overlap in PWI tasks results in

little or no priming (e.g., Wong & Chen, 2008, 2009) hence we had to define overlap

syllabically (here, in terms of the initial syllable overlapping between target and distractor).

It is worth highlighting that the relative degree of phonological overlap between target and

distractor was comparable across the two experiments: average segmental overlap was

54% in Experiment 1 (see "Materials" section), and syllabic overlap was 50% in Experiment

2 (one out of two syllables was shared between disyllabic targets and distractors).

What are the theoretical implications of the current findings with regard to models of

phonological encoding across languages? Due to the sparsity of available results on word

production in non-Western languages, theoretical accounts of the results from our

experiments are necessarily speculative. Figure 4 shows a rough processing sketch of

word form encoding across the two target languages, loosely adapted from Roelofs (2015)

and O'Seaghdha (2015). "Proximate units", defined as the primary selectable unit below

the word level by O'Seaghdha et al. (2010) are highlighted. Note that the Mandarin model

contains a segmental layer, despite the fact that behavioral experiments which manipulated

segmental overlap have tended to result in null findings (e.g., Chen et al., 2002; Wong &

Chen, 2008, 2009; see also Verdonschot et al., 2012, for results from Japanese).

Segments nevertheless probably contribute to phonological encoding because segmentally

based speech errors are found in spoken Chinese (e.g., Chen, 1993); note also that Qu,

Damian, & Kazanina (2012; see also Yu, Mo & Mo, 2014) presented EEG evidence for the

presence of segmental effects in Mandarin speakers despite behavioral null findings, which

further warrants inclusion of such a segmental layer in the model.

Given that phonological overlap in our experiments was defined at the proximate unit

level (i.e., segmental in English but based on atonal syllables in Mandarin) there is no

obvious reason why information transmission from morpheme to proximate unit level

should be cascaded in English, but serial in Mandarin. Nevertheless, this is what our results

suggest. Hence, processing along the critical pathway (shown in dotted lines in Figure 4)

could fundamentally differ between languages. This account is admittedly post hoc, and it

is not obvious why this should be the case. Perhaps this is because in Western languages,

relatively few segments combine to form a potentially unlimited number of lexical items,

whereas in Mandarin, the number of syllables is much larger and so a discrete activation

makes sense for syllables whereas the process is more continuous for segments.

(Insert Figure 4 here)

In all three relevant sets of results (Zhu et al., 2015, 2016; and the current findings) phonological overlap was defined in terms of overlapping atonal syllables (e.g., the target "cherry", /ying1tao2/ was paired with the distractor "profit", /ying2li4/). Given that in PWI tasks, segmental overlap in (Cantonese) Chinese by itself does not generate priming (Wong & Chen, 2008, 2009), we attributed the phonological facilitation in our Experiment 2 to the "proximate unit" layer of atonal syllables. Is it possible that this assumption is incorrect and phonological facilitation perhaps took place at the level of the tonally specified syllable? If so, this could explain the statistical additivity between semantic interference and phonological facilitation reported in Zhu et al. (2016), as well as the relatively late time window under which phonological effects appeared in the EEG results reported by Zhu et al. (2015). Such an account would have to explain why the co-activation of similar (but tonally mismatching) syllables generates behavioral priming. Under the assumption that segments and tonal syllables are bidirectionally connected, perhaps the activation of the distractor syllable /ying2/ could prime, via shared segments, the target syllable /ying1/.

Although this scenario is not impossible, we believe that similar effects would then be predicted for Western languages as well: e.g., if both the target object "cat" and a phonologically related distractor "cap" activate their corresponding syllables and activation is shared between them via segmental overlap, then we should also find syllabic priming in English and other Western languages. However, this is not the case (e.g. Schiller, 1998, 2000; Schiller, Costa & Colomé, 2002), nor does the prediction agree with EEG studies of spoken word production (see Introduction) which have shown a "late" time window of

phonological effects only in Mandarin, but not in Western languages. Furthermore, it is interesting to note that the WEAVER model of word form encoding (Levelt et al., 1999; Roelofs, 1997) stipulates that access to syllable program nodes is competitive. Hence the prediction from this framework is that co-activation of similar syllables should hinder, rather than facilitate, access to the correct target syllable. This clearly conflicts with the fact that phonological overlap in PWI generally results in facilitation.

Clearly, further research is required to resolve this issue, and we acknowledge the need for alternative approaches concerning how the phonological properties of a target language could affect semantic-to-phonological transmission. Especially, other phonological properties such as neighborhood density (Peramunage, Blumstein, Myers, Goldrick, & Baese-Berk, 2011), the role of the relatively low number of atonal syllables and unique properties of tones in Chinese (Roelofs, 2015), need to be investigated directly in the future.

To summarize, the results of the present study suggest that in English spoken word production, non-target lexical entries can under certain circumstances activate their corresponding phonological properties, hence supporting a notion of lexical access in which information transmission from semantic to phonological layers is cascaded. This pattern dovetails with a rising number of findings from various tasks and conducted with speakers of Western languages. By contrast, in Mandarin word production, no such evidence for cascadedness was found, and information transmission appeared more strictly serial. This discrepancy highlights potential fundamental differences in phonological encoding across target languages. The combination of semantic blocking and picture-word

interference used here and previously (Aristei et al., 2011) offers new possibilities for

investigating the underlying mechanisms of spoken word production.

.

**References**

Abdel Rahman, R., & Melinger, A. (2007). When bees hamper the production of honey:

Lexical interference from associates in speech production. *Journal of Experimental*

*Psychology: Learning, Memory, and Cognition, 33*, 604–614.

Abdel Rahman, R., & Melinger, A. (2008). Enhanced phonological facilitation and traces of

concurrent word form activation in speech production: An object-naming study with

multiple distractors. *The Quarterly Journal of Experimental Psychology, 61*,

1410–1440.

Aristei, S., Melinger, A., Abdel Rahman, R. (2011). Electrophysiological chronometry of

semantic context effects in language production. *Journal of Cognitive Neuroscience,*

*23*, 1567–1586.

Belke, E., Meyer, A. S., & Damian, M. F. (2005). Refractory effects in picture naming as

assessed in a semantic blocking paradigm. *Quarterly Journal of Experimental*

*Psychology: Human Experimental Psychology, 58*, 667–692.

Bonin, P., & Fayol, M. (2000). Writing words from pictures: What representations are

activated, and when? *Memory & Cognition, 28*, 677–689.

Chen, J.-Y. (1993). A small corpus of speech errors in Mandarin Chinese and their

classification. *The World of Chinese Language, 69*, 26–41 (in Chinese).

Chen, J.-Y., Chen, T. -M., & Dell, G. S. (2002). Word-form encoding in Mandarin Chinese

as assessed by the implicit priming task. *Journal of Memory and Language, 46*,

751–781.

Chen, J.-Y., Lin, W. -C., & Ferrand, L. (2003). Masked priming of the syllable in Mandarin

Chinese syllable production. *Chinese Journal of Psychology, 45*, 107–120.

Chinese Linguistic Data Consortium. (2003). 现代汉语通用词表[Chinese lexicon]

(CLDC-LAC-2003-001). Beijing, China: Tsinghua University, State Key Laboratory of

Intelligent Technology and Systems, and Chinese Academy of Sciences, Institute of

Automation.

Costa, A., Strijkers, K., Martin, C., & Thierry, G. (2009). The time course of word retrieval

revealed by event-related brain potentials during overt speech. *Proceedings of the*

*National Academy of Sciences of the United States of America, 106(50),*

21442-21446.

Damian, M. F. (1998). *Semantic and phonological factors in speech production: Evidence*

*from picture-word interference experiments.* Unpublished doctoral dissertation, Rice

University, Houston, USA.

Damian, M. F., & Als, L. C. (2005). Long-lasting semantic context effects in the spoken

production of object names. *Journal of Experimental Psychology: Learning, Memory,*

*and Cognition, 31*(6), 1372–1384.

Damian, M. F., & Bowers, J. S. (2009). Assessing the role of orthography in speech

perception and production: Evidence from picture-word interference tasks. European

Journal of Cognitive Psychology, 21(4), 581-598.

Damian, M. F., & Martin, R. C. (1999). Semantic and phonological codes interact in single

word production. *Journal of Experimental Psychology: Learning, Memory, and*

*Cognition, 25*(2), 345–361.

Damian, M. F., Vigliocco, G., & Levelt, W. J. M. (2001). Effects of semantic context in the

naming of pictures and words. *Cognition, 81*, B77–B86.

Davis, C. J. (2005). N-Watch: A program for deriving neighborhood size and other

psycholinguistic statistics. *Behavior Research Methods, 37*, 65–70.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production.

*Psychological Review, 93*(3), 283–321.

Dell, G. S., & Reich, P. A. (1981). Stages in sentence production: An analysis of speech

error data. *Journal of Verbal Learning and Verbal Behavior, 20*, 611-629.

Dell'Acqua, R., Sessa, P., Peressotti, F., Mulatti, C., Navarrete, E., & Grainger, J. (2010).

ERP evidence for ultra-fast semantic processing in the picture-word interference

paradigm. *Frontiers in Psychology,* 1:177.

Dumay N., Damian M. F. (2011). A word-order constraint in single-word production? Failure

to replicate Janssen, Alario, and Caramazza (2008). *Psychological Science, 22*,

559–561.

Forster, K. I., & Forster, J. C. (2003). DMDX: A windows display program with millisecond

accuracy. *Behavior Research Methods, Instruments and Computers, 35*, 116–124.

Glaser, W. R., & Düngelhoff, F. J. (1984). The time course of picture word interference.

*Journal of Experimental Psychology: Human Perception and Performance, 10*(5),

640–654.

Humphreys, G. W., Riddoch, M. J., & Quinlan, P. T. (1988). Cascade processes in picture

identification. *Cognitive Neuropsychology, 5*(1), 67–104.

Jescheniak, J. D., Hahne, A., & Schriefers, H. (2003). Information flow in the mental lexicon:

Evidence from event-related brain potentials. *Cognitive Brain Research, 15*,

261–276.

Jescheniak, J. D., Hahne, A., Hoffmann, S., & Wagner, V. (2006). Phonological activation

of category coordinates during speech planning is observable in children but not in

adults: Evidence for cascaded processing. *Journal of Experimental Psychology:*

*Learning, Memory, and Cognition, 32*, 373–386.

Jescheniak, J. D., & Levelt, W. J. M. (1994). Word frequency effects in speech production:

Retrieval of syntactic information and of phonological form. *Journal of Experimental*

*Psychology: Learning, Memory, and Cognition, 20,* 824–843.

Jescheniak, J. D., & Schriefers, H. (1998). Discrete serial versus cascaded processing in

lexical access in speech production: Further evidence from the coactivation of

near-synonyms. *Journal of Experimental Psychology: Learning, Memory, and*

*Cognition, 24*, 1256–1274.

Kuipers, J., & La Heij, W. (2009). The limitations of cascading in the speech production

system. *Language and Cognitive Processes, 24*, 120–135.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech

production. *Behavioral and Brain Sciences, 22*(1), 1–75.

Love, J., Selker, R., Marsman, M., Jamil, T., Dropmann, D., Verhagen, A. J., Ly, A., Gronau,

Q. F., Smira, M., Epskamp, S., Matzke, D., Wild, A., Rouder, J. N., Morey, R. D. &

Wagenmakers, E.-J. (2015). JASP (Version 0.7)[Computer software].

Maess, B., Friederici, A. D., Damian, M., Meyer, A. S., & Levelt, W. J. (2002). Semantic

category interference in overt picture naming: Sharpening current density localization

by PCA. *Journal of Cognitive Neuroscience,14*(3), 455–462.

Melinger, A., & Abdel Rahman, R. (2013). Lexical selection is competitive: Evidence from

    indirectly activated semantic associates during picture naming. *Journal of*

    *Experimental Psychology: Learning, Memory, and Cognition, 39*(2), 348–364

Meyer, A. S., & Damian, M. F. (2007). Activation of distractor names in the picture-picture

    interference paradigm. *Memory and Cognition, 35*, 494–503.

Morsella, E., & Miozzo, M. (2002). Evidence for a cascade model of lexical access in

    speech production. *Journal of Experimental Psychology: Learning, Memory, and*

    *Cognition, 28,* 555–563.

Navarrete, E., & Costa, A. (2005). Phonological activation of ignored pictures: Further

    evidence for a cascade model of lexical access. *Journal of Memory and Language,*

    *53*, 359–377.

O'Seaghdha, P. G. (2015). Across the great divide: Proximate units at the

    lexical-phonological interface. *Japanese Psychological Research, 57*, 4–21.

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.

Oppenheim, G. M., Dell, G. S., & Schwartz, M. F. (2010). The dark side of incremental

    learning: A model of cumulative semantic interference during lexical access in

    speech production. *Cognition, 114*, 227–252.

Oppermann, F., Jescheniak, J. D., Schriefers, H., & Görges, F. (2010). Semantic

    relatedness among objects promotes the activation of multiple phonological codes

    during object naming. *The Quarterly Journal of Experimental Psychology, 63*,

356–370.

Peramunage, D., Blumstein, S. E., Myers, E. B., Goldrick, M., & Baese-Berk, M. (2011).

Phonological neighborhood effects in spoken word production: An fMRI

study. *Journal of Cognitive Neuroscience*, *23*(3), 593–603.

Peterson, R. R., & Savoy, P. (1998). Lexical selection and phonological encoding during

language production: Evidence for cascaded processing. *Journal of Experimental

Psychology: Learning, Memory, and Cognition, 24*, 539–557.

Qu, Q. Q., Damian, M. F., & Kazanina, N. (2012) Sound-sized segments are significant for

Mandarin speakers. *Proceedings of the National Academy of Sciences of the United

States of America, 109*(35):14265–14270.

Raftery, A. E. (1999). Bayes factors and BIC: Comment on "A critique of the Bayesian

information criterion for model selection". *Sociological Methods & Research, 27*,

411–427.

Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition,

42*, 107–142.

Roelofs, A. (1997). The WEAVER model of word-form encoding in speech production.

*Cognition, 64,* 249–284.

Roelofs, A. (2008). Tracing attention and the activation flow in spoken word planning using

eye movements. *Journal of Experimental Psychology: Learning, Memory, and

Cognition, 34,* 353–368.

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.

Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes

factors for ANOVA designs. *Journal of Mathematical Psychology, 56*(5), 356–374.

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t

tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin &*

*Review, 16*(2), 225–237.

Roux, S., Bonin, P., & Kandel, S. (2014). The "BIG BIRD" of the "YELLOW YOUNG" man:

Do nontarget properties cascade? *Quarterly Journal of Experimental Psychology, 67*,

763–784.

Schiller, N. O. (1998). The effect of visually masked syllable primes on the naming

latencies of words and pictures. *Journal of Memory and Language, 39*, 484-507.

Schiller, N. O. (2000). Single word production in english: the role of subsyllabic units during

phonological encoding. *Journal of Experimental Psychology: Learning, Memory, and*

*Cognition, 26*, 512-521,

Schiller, N. O., Costa, A., & Colomé, A. (2002). Phonological encoding of single words: In

search of the lost syllable. *Laboratory Phonology, 7*, 35-59.

Schriefers, H., Meyer, A. S., & Levelt, W. J. M. (1990). Exploring the time course of lexical

access in language production: Picture-word interference studies. *Journal of Memory*

*and Language, 29*(1), 86–102.

Snodgrass, J. C., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for

names agreement, image agreement, familiarity, and visual complexity. *Journal of*

*Experimental Psychology: Human Learning and Memory, 6*, 174–215.

Starreveld, P. A., & La Heij, W. (1995). Semantic interference, orthographic facilitation, and

their interaction in naming tasks. *Journal of Experimental Psychology: Learning,*

*Memory, and Cognition, 21*(3), 686–698.

Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method.

*Acta Psychologica, 30*, 276–315.

Taylor, J. K., & Burke, D. M. (2002). Asymmetric aging effects on semantic and

phonological processes: Naming in the picture-word interference task. *Psychology*

*and Aging, 17*(4), 662–676.

Verdonschot, R. G., Nakayama, M., Zhang, Q., Tamaoka, K., & Schiller, N. O. (2013). The

proximate unit of Chinese-English bilinguals: Proficiency matters. *PLoS ONE,*

*8(4)*:e61454. Doi:10.1371/journal.pone.0061454.

Verdonschot, R. G., Kiyama, S., Tamaoka, K., Kinoshita, S., La Heij, et al. (2011). The

functional unit of Japanese word naming: Evidence from masked priming. J*ournal of*

*Experimental Psychology: Learning, Memory, and Cognition, 37,* 1458-1473.

Wong, A. W. K., & Chen, H. -C. (2008). Processing segmental and prosodic information in

Cantonese word production. *Journal of Experimental Psychology: Learning, Memory,*

*and Cognition, 34,* 1172–1190.

Wong, A. W.-K., & Chen, H.-C. (2009). What are effective phonological units in Cantonese

spoken word planning? *Psychonomic Bulletin and Review, 16*, 888–892.

Yu, M., Mo, C., & Mo, L. (2014). The role of phoneme in Mandarin Chinese production:

Evidence from ERPs. *PLoS ONE 9(9): e106486.*

You, W., Zhang, Q., & Verdonschot, R. G. (2012). Masked syllable priming effects in word

and picture naming in Chinese. *PloS ONE, 7*(10), e46595.

Zhang, Q., Chen, H.-C., Weekes, B. S., & Yang, Y. (2009). Independent effects of orthographic and phonological facilitation on spoken word production in Mandarin. *Language Speech, 52*(1), 113–126.

Zhang, Q., & Yang, Y. (2003). The determiners of picture naming latency. *Acta Psychologica Sinica, 35,* 447–454 (in Chinese).

Zhang, Q., & Yang, Y. (2005). The phonological planning unit in Chinese monosyllabic word production. *Psychological Science, 28*, 374–378 (in Chinese).

Zhao, H., La Heij, W., & Schiller, N. O. (2012). Orthographic and phonological facilitation in speech production: New evidence from picture naming in Chinese. *Acta Psychologica*, 139, 272–280.

Zhu, X., Zhang, Q., & Damian, M. F. (2016). Additivity of semantic and phonological effects: Evidence from speech production in Mandarin. *Quarterly Journal of Experimental Psychology.* DOI: 10.1080/17470218.2015.1129427

Zhu, X., Damian, M. F., & Zhang, Q. (2015). Seriality of semantic and phonological processes during overt speech in Mandarin as revealed by event-related brain potentials. *Brain and Language, 144,* 16–25.

**Footnotes**

[1]In a number of previous studies (e.g., Aristei et al., 2011; Belke et al., 2005; Damian & Als, 2005) the effect of repetition/cycle was explicitly included in the experimental design. However, in the current experiments the design is already reasonably complex, and in the interest of clarity we limited our analyses to those most economically designed to assess our predictions.

[2]We thank an anonymous reviewer for bringing this point to our attention.

[3]Again, we thank an anonymous reviewer for bringing up this issue.

**Figure Captions**

*Figure 1.* Picture-word interfence task combined with semantic blocking: Sample trials from a semantically homogeneous block (category: body parts) with unrelated, phonologically related, and mediated distractor words.

*Figure 2.* Mean pictures naming latencies and standard errors dependent on semantic context and distractor type in English. *$p < .05$; **$p<.01$; ***$p<.001$.

*Figure 3.* Mean pictures naming latencies and standard errors dependent on semantic context and distractor type in Mandarin. *$p < .05$; **$p<.01$; ***$p<.001$.

*Figure 4.* Sketch of phonological encoding in English and Mandarin. Proximate units are highlighted.
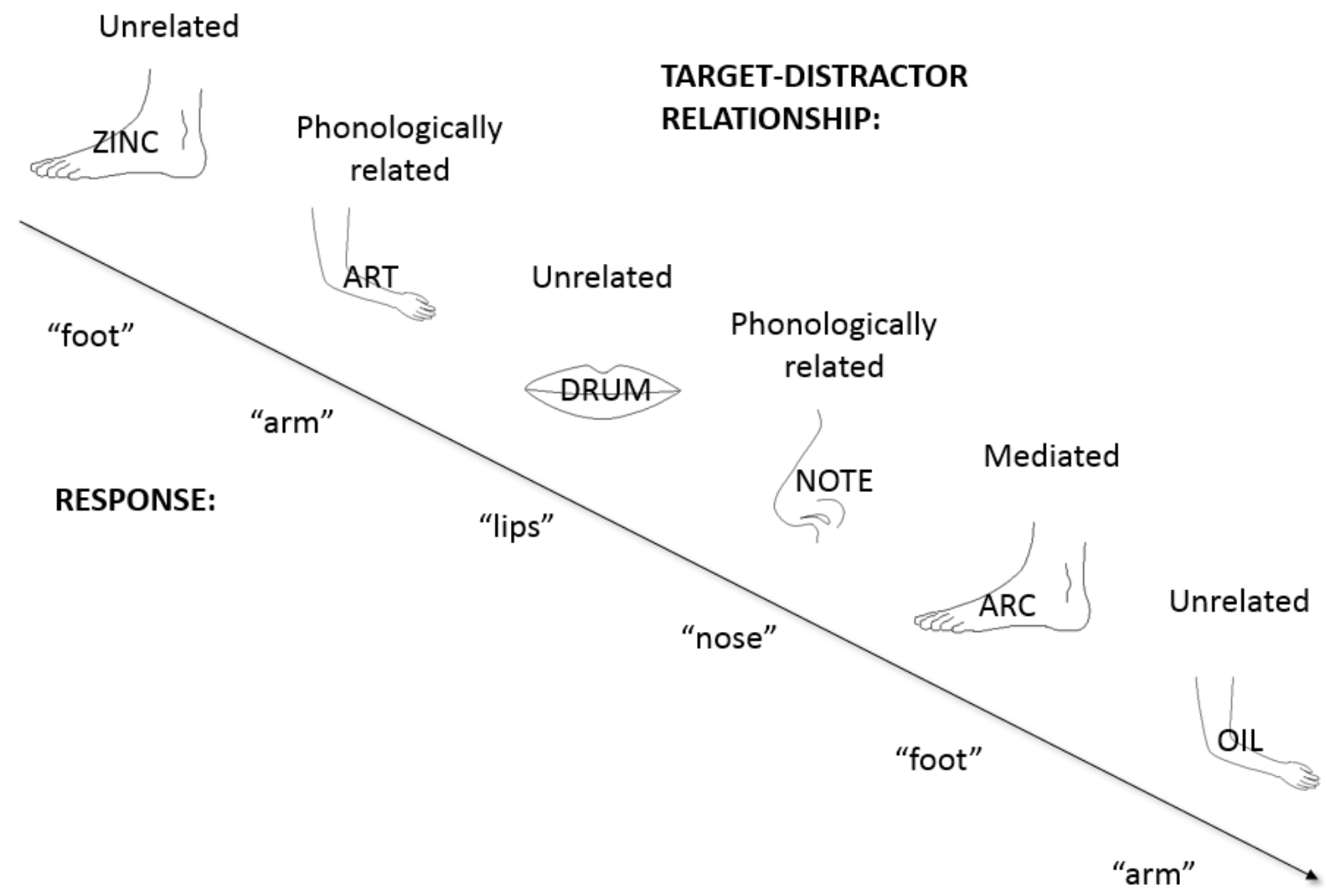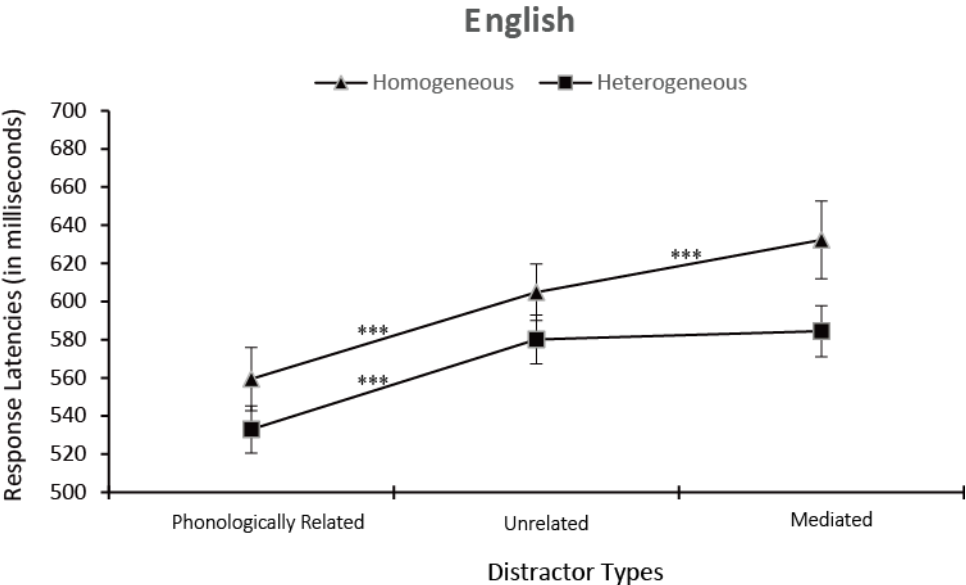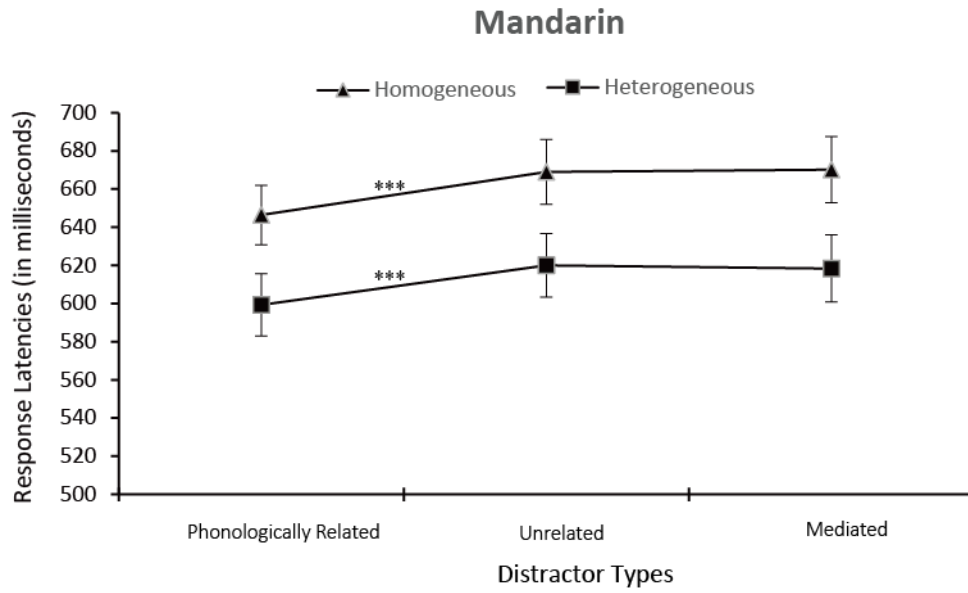
Figure 1

Unrelated

ZINC

Phonologically
related

**TARGET-DISTRACTOR
RELATIONSHIP:**

ART

Unrelated

"foot"

Phonologically
related

DRUM

"arm"

NOTE

Mediated

**RESPONSE:**

"lips"

ARC

Unrelated

"nose"

"foot"

OIL

"arm"

Figure 2



English

Homogeneous    Heterogeneous

Figure 3

Figure 4

**ENGLISH**                                **MANDARIN**