Mem Cogn (2015) 43:111–120
DOI 10.3758/s13421-014-0451-9

Effects in production of word pre-activation during listening: Are listener-generated predictions specified at a speech-sound level?

Eleanor Drake · Martin Corley

Abstract It has been demonstrated that listener-generated predictions of upcoming material can be specified to a phonological level, such that a specific word onset is anticipated (e.g., DeLong, Urbach, & Kutas, *Nature Neuroscience*, *8*, 1117–1121, 2005). In the present study, we investigated whether such word-form-specific predictions impact picture-naming latencies in a manner similar to that observed when a distractor word is actually presented. Participants were auditorily presented with high-cloze sentence stems, in order to elicit word-form predictions. The pictures for naming were presented immediately following the sentence stem. We systematically manipulated the phonological relationship between the predicted word and the picture name. Across three experiments, naming was facilitated when the picture name fully matched the predicted word. However, naming was neither facilitated nor inhibited when the picture name overlapped phonologically with the predicted word. This finding is in contrast to the known effects of phonological overlap when a distractor word is heard or read. Our findings suggest that words that are internally listener-generated (predicted) during comprehension are not robustly specified at a speech-sound (phonological) level.

Keywords Language comprehension · Language production · Word production · Word prediction

The speech production system is active during comprehension. Perceiving speech primes the speech motor system, even when no speech output is required. Articulatory muscles are activated when listening to speech sounds, but not when listening to nonspeech sounds, and such increased excitability of the motor system is accompanied by an increase in activity within Broca's area (Fadiga, Craighero, Buccino, & Rizzolatti, 2002; Watkins, Strafella, & Paus, 2003; see also Pulvermüller et al., 2006). Activation of speech–motor areas during speech comprehension may reflect backward- or forward-looking processes, or both. The production system may be engaged in generating articulatory representations to support the maintenance and decoding of heard material; equally, it may be engaged in simulating upcoming auditory input via the generation of emulations (Pickering & Garrod, 2004; Watkins & Paus, 2004). Recently, there has been an increased focus on the latter possibility. It has influentially been suggested that during comprehension the production system is engaged in generating predictions of upcoming material, thereby reducing processing demands on the comprehension system by constraining the possible interpretations of incoming material (e.g., Pickering & Garrod, 2007; Schiller, Horemans, Ganushchak, & Koester, 2009; see also Scott, McGettigan, & Eisner, 2009, for a review). But whereas evidence of speech–motor activation during comprehension is compatible with the notion that the listener's speech–motor system is engaged in generating predictions, it does not constitute proof. As a first step toward confirming such an interpretation, it would be necessary to demonstrate that upcoming material in specific is represented via the speech production system to at least a speech-sound level (see Hickok, 2012).

Although it has not been empirically demonstrated that predictions during comprehension are effected via the speech–motor areas, it has been demonstrated that words are predicted at a surface-form level during comprehension (DeLong, Urbach, & Kutas, 2005). When reading sentences

E. Drake · M. Corley (✉)
Psychology, PPLS, University of Edinburgh, 7 George Square,
Edinburgh EH8 9JZ, UK
e-mail: Martin.Corley@ed.ac.uk

that strongly predict a noun (such as *The day was breezy so the boy went outside to fly . . .*), comprehenders exhibit increased N400 amplitudes upon encountering an indefinite article in a form inappropriate to the predicted noun (e.g., *an*, where the prediction is *kite*). The amplitude of this response is correlated with the probability that the predicted noun completes the sentence, as determined by previous offline testing. This effect can relate only to the upcoming word's being specified at a phonological-form level, because the distinction between *a* versus *an* is empty at a semantic and syntactic level, and is based purely on the phonological form of the upcoming word (consonant vs. vowel). It is not clear whether the phonological form representation that is elicited encompasses activation of a speech-sound-level representation within the speech production system of the listener.

In the present study, we investigated whether effects of phonological-form prediction in comprehension are observable in a behavioral measure of speech production. If prediction does involve the speech production system in the generation of speech-sound emulations, we would expect to see an effect of prediction in the listener's own speech production system. Previous findings from picture–word interference, picture–picture interference, and sentence-listening paradigms provide some guidance as to the possible nature of such an effect.

Picture-naming is facilitated when the picture is accompanied by a partially phonologically overlapping written distractor word (as compared to one with no overlap; Damian & Dumay, 2007; Lupker, 1982). When the distractor word is presented auditorily, picture naming is facilitated only when there is onset overlap (Meyer & Schriefers, 1991). Overall, pictures are named more slowly in the presence of a distractor word than in isolation (Meyer & Schriefers, 1991). This pattern of findings may result from the production system being habitually and automatically recruited during comprehension, causing presentation of a distractor word to increase the demands on the production system, and thereby leading to a general increase in picture-naming latencies (see Greene, 1988; Raney, 1993). If so, the phonological facilitation effects described above may be better understood as attenuated inhibition effects, in which the inhibition is attenuated as a consequence of the overlap between competing representations in the production system.

In typical picture–word interference experiments, the distractor is orthographically represented. In studies in which the distractor is a picture, participants must internally generate the lexical form of the distractor, and effects of phonological overlap are found in some studies (e.g., Morsella & Miozzo, 2002; Navarette & Costa, 2005), but not others (Jescheniak et al., 2009). This difference has recently been attributed to aspects of the distractor pictures used (Oppermann, Jescheniak, & Görges, 2014), potentially rendering comparison with distractors that are (implicitly) predicted during comprehension difficult. However, one study in which target (rather than distractor) names were internally generated provides evidence that such word forms can be phonologically specified in a way that affects speech production latencies (Humphreys, Boyd, & Watter, 2010). In a free-association version of the picture–word interference paradigm, each written word had a single, high-likelihood associate (e.g., "cobweb" → spider). Participants named the first word that came to mind as an associate of the written word, while ignoring the picture. In phonologically related trials, the associate and the picture name shared an onset (e.g., "cobweb" → *sp*ider; *SP*OON), whereas in unrelated trials there was no such phonological overlap (e.g., "cobweb" → *sp*ider; *F*ORK). Response latencies were significantly shorter in the phonologically related condition than in the unrelated condition, and did not differ from those in the control condition[1] (in which participants did not see any picture). This suggests that there is no need for a word to be perceptually available in order to elicit effects at a phonological level, and strengthens the case for suggesting that the locus of any facilitation is in the production system. If prediction during comprehension is production-driven, then we might expect words predicted during comprehension to elicit phonological effects similar to those of pictures implicitly named during viewing.

Previous studies concerning the effect of sentence-stem context on picture naming have investigated integration rather than prediction effects, and have therefore focused on the effects of manipulating the semantic and/or syntactic congruence between the sentence stem and the picture name (e.g., Roe et al., 2000; see also Griffin & Bock, 1998; Wicha et al., 2005). Pictures are named fastest in a congruent context, more slowly in a neutral context, and most slowly in an incongruent context. This pattern has been interpreted as indicating (prediction-mediated) easing of integration (Griffin & Bock, 1998; Wicha et al., 2005). Of course, the effect is also consistent with a prediction-as-production account. This interpretation merits further exploration, particularly in light of evidence that inhibition of picture-naming may arise from conflict at the production level, rather than from integration costs (Hirschfeld, Jansma, Bölte, & Zwitserlood, 2008; Mahon, Costa, Peterson, Vargas, & Caramazza, 2007; Nozari, Dell, & Schwartz, 2011; Severens, Janssens, Kühn, Brass, & Hartsuiker, 2011).

In summary, the speech production system is active during comprehension, and comprehension can involve prediction of word forms at a phonological level. These findings have been linked in the suggestion that prediction during comprehension engages the speech production system. The present study was designed to explore this suggestion directly. In three experiments, participants heard auditory sentence fragments with

---

[1] The statistic for related versus control conditions was not reported, and this conclusion is drawn from Table 1 and Fig. 2 of Humphreys et al. (2010).

highly predictable continuations (such as *He managed to fix the drip from the old leaky*. . .) and named pictures at the offset of the audio. Picture names were chosen such that they corresponded to the predictable word (tap–*TAP*) or had a partial phonological overlap (tap–*CAP*, tap–*TAN*), or had no overlap (tap–*CONE*). In order to maximize the probability of phonological effects, each mismatching sentence continuation corresponded to an image name on other trials (cf. Meyer & Damian, 2007). We predicted facilitation of picture-naming in the matching condition (as compared to an acontextual control condition), as evidence that participants were making predictions as a consequence of hearing the sentence stems. Of interest was whether we would see an effect of phonological overlap, which would confirm that during speech comprehension, predicted items are activated at a phonological level in the production system.

Method

Participants

Twenty-seven adults (17 female, ten male) with a mean age of 19 years (range 18–24) participated in Experiment 1. A total of 21 further adults (14 female, seven male) with a mean age of 20 years (range 18–26) participated in Experiment 2. Finally, a further 21 adults (14 female, seven male) with a mean age of 20 years (range 18–27) participated in Experiment 3. The participants were students from the University of Edinburgh, who either received course credit or were paid for participation. One participant in each of Experiments 1 and 2 identified themselves as multilingual subsequent to the recordings; the data from these participants were excluded from the analyses. All remaining participants were monolingual speakers of English. No participant reported relevant language or visual impairments. Written consent was provided, in line with British Psychological Society guidelines.

Materials

We used identical sets of sentence stems and of pictures in three experiments. Sentence stems were chosen such that they strongly predicted the following word; depending on the condition in which they were encountered, pictures had names that either corresponded to the predictable word, overlapped with it phonologically, or were unrelated.

Twelve pictures were used as experimental items. A further 12 pictures acted as filler items, included in order to minimize participants' conscious attention to phonological aspects of the picture names, and to maintain an even balance between trials in which pictures corresponded to predicted or unpredicted names. Pictures were pretested online for name agreement (median agreement = .85, mean = .76; range = .3 to 1), and their names were all monosyllabic concrete nouns of medium frequency (mean $\log_{10}$CD = 2.93, *SD* = 0.41, range = 2.07–3.91; SUBTLEX-US database, Brysbaert & New, 2009). We used 36 experimental sentence stems, three of which predicted the name of each of the 12 experimental pictures, as determined by pretest (all cloze likelihoods > .8). Similarly, three of 36 filler sentence stems predicted the name of each of the 12 filler pictures. Sentence stems were recorded by a female native speaker of British English. In experimental trials, we manipulated the phonological relationship between the word predicted by the sentence stem, and the accompanying picture presented for naming. In Experiment 1, we included matching (tap–*TAP*), onset-overlap (tap–*TAN*), and rime-overlap (tap–*CAP*) conditions. We compared naming latencies in these conditions to those in a control condition in which the picture was named following backward speech (maintaining the "speech-like" qualities of the sentence stems, but ensuring that there were no linguistic cues). In Experiment 2, we replaced the onset-overlap condition with a no-overlap condition (tap–*CONE*). In Experiment 3, we replaced the rime-overlap condition of Experiment 2 with an onset-overlap condition, allowing for direct comparison of the onset-overlap and no-overlap conditions, which differed in Humphreys et al.'s (2010) word association study. In each experiment, all experimental pictures occurred in all conditions for all participants, allowing each picture to act as its own control in a fully within-participants and within-items design.

Procedure

The experiments were presented using DMdX (Forster & Forster, 2003). Each experiment began with a familiarization phase, during which participants saw each of the 24 experimental and filler pictures accompanied by a printed name, and named the picture aloud. Each picture and corresponding name appeared three times in total. In each experiment, the familiarization procedure was followed by five experimental blocks. Blocks 1 and 5 were included principally in order to provide a control condition for a parallel speech-imaging study. For the purposes of the present study, these blocks allowed us to confirm that participants were able to correctly name all pictures when these were presented in isolation. Each picture was presented on its own and named aloud. Participants viewed a fixation point in the center of the screen for 2.9 s (which was the mean duration of the sentence-stem recordings used in other blocks), immediately prior to the presentation of the picture-to-be-named. Participants were instructed to name each picture as quickly and accurately as possible (as they had practiced during the familiarization phase). In Blocks 2, 3, and 4, participants again viewed a fixation point, but this time while listening to a sentence stem.

The picture to be named was presented immediately at the offset of the last word of the sentence stem. Participants were again instructed to name pictures as quickly and accurately as they could. In each of Blocks 2, 3, and 4, each experimental picture was presented four times: once in each experimental condition, and once in the backward-speech control condition. Each sentence stem was heard by each participant once per block, and once per condition across the three blocks. We manipulated the condition in which a picture was encountered by altering the sentence stem with which it was presented. Item presentation within each block was fully randomized so that all conditions were interleaved with one another and with filler items.

Picture-naming response latencies were automatically recorded, together with full auditory responses, by the experimental software. We also made an independent audio-recording of each full session. Participants took no more than 50 min to complete an experiment.

Results

All analyses were conducted using the lme4 package, version 0.999999-4, in R 3.0.3 (Bates, Maechler, & Bolker, 2013; R Development Core Team, 2014). For response times, we used linear mixed-effects models fit by maximum likelihood; error responses were analyzed using binomial mixed-effects models, fit using Laplace estimation. Linear models included a predictor of *log trial position* to account for practice effects throughout the experiment (due to low numbers of errors, binomial models did not include this predictor). In each case, we included the effects of *context* (matching, rime-overlap, and onset-overlap [Exp. 1]; matching, rime-overlap, and no-overlap [Exp. 2]; matching, onset-overlap, and no-overlap [Exp. 3]; and backward speech as a control) on the response variable of interest. The *context* predictor was orthogonally coded, as detailed below for each experiment. Following suggestions made by Barr, Levy, Scheepers, and Tily (2013), each model was "maximally specified," with both intercepts and slopes, as well as their correlations, being allowed to vary by participants and, where possible, by items. In linear analyses, we treated *t*s of 2.00 and above as significant, due to complexities in estimating the degrees of freedom associated with predictors (Baayen, 2008).

Experiment 1

In Experiment 1 there was a total of 3,744 recorded responses, of which 65 (1.7 %) were errors, either because of lexical intrusion of the word predicted by the context, or because of other factors (such as failure to record a response). Table 1 provides a summary of error numbers by conditions; Fig. 1 shows the by-participants mean times to respond correctly in each condition. We performed two analyses: One of latencies to (correctly) name the pictures presented, and one of the likelihood of producing an error. In each analysis, we used orthogonal contrasts for the *context* predictor, such that we first compared observations for the match condition to those for all other conditions; second, we compared the two conditions in which a word different from the picture was predicted by the context (mismatch conditions) to the backward control condition; and third, we compared the two mismatch conditions (rime overlap vs. onset overlap).

The model of response times included *log trial position* and *context* as predictors, as well as fixed and random intercepts, with random effects of log trial position and context by participants, and of context by items. The fixed effect of context significantly improved the model fit [$\chi^2(3) = 29.5$, $p < .001$]. Model coefficients showed that participants' responses became faster as the experiment progressed [$\beta = –11.7$, $SE(\beta) = 4.8$, $t = 2.44$] and that participants were faster to correctly name pictures in the match condition than in the other three conditions [$\beta = –18.2$, $SE(\beta) = 2.2$, $t = 8.26$]. We found no differences within the three nonmatching conditions (*t*s < 1). Table 2 gives details of the model coefficients.

To avoid empty cells, we analyzed total error numbers, rather than contextual errors alone. Because of the low numbers of recorded errors of any type, the analysis of errors did not include trial position; nor did it include a random effect of context by items, due to a failure of the model to converge. Including a fixed effect of context significantly improved the model fit [$\chi^2(3) = 23.8$, $p < .001$].The coefficients showed significantly fewer errors in the match condition than in the other three conditions combined [$\beta = –0.62$, $SE(\beta) = 0.15$, $z = –4.06$, $p < .001$] and more errors in the two overlap conditions than in the backward control condition [$\beta = 0.87$, $SE(\beta) = 0.19$, $z = 4.51$, $p < .001$]. No difference was apparent between the two overlap conditions ($z < 1$). Table 3 gives details of the model coefficients.

Experiment 2

In Experiment 2, 51 of 2,879 responses (1.8 %) were errors. Errors and picture-naming latencies are summarized in Table 1 and in Fig. 1, respectively. In the analyses of response times and of error rates, we used orthogonal contrasts similar to those for Experiment 1, such that the matching condition was compared to all others, the two conditions with mismatching predictions were compared to the backward control condition, and finally, these two conditions (rime overlap vs. no overlap) were compared to each other.

Other than the differences in conditions, the details of the model construction were identical to those for Experiment 1. The fixed effect of context again improved the model fit [$\chi^2(3) = 27.4$, $p < .001$]. Participants again responded faster

Table 1 Recorded errors in Experiments 1–3

	Context				
Experiment	Match	Rime Overlap	Onset Overlap	No Overlap	Backward
1	5/– (0.4 %)	23/18 (2.5 %/1.9 %)	29/24 (3.1 %/2.6 %)	–	8/– (0.5 %)
2	4/– (0.6 %)	23/20 (3.1 %/2.8 %)	–	16/14 (2.2 %/1.9 %)	8/– (1.1 %)
3	1/– (0.1 %)	–	18/17 (2.7 %/2.6 %)	12/10 (1.8 %/1.5 %)	6/– (0.9 %)

Numbers refer to the total errors/numbers of errors in which a distractor was produced in error (percentages in parentheses)

as the experiment progressed [$\beta = -15.8$, $SE(\beta) = 4.0$, $t = 3.96$] and were fastest to respond in the match condition relative to the other conditions [$\beta = -19.7$, $SE(\beta) = 2.4$, $t = 8.33$]. No differences emerged within the three nonmatching conditions (*t*s < 1.29). The model coefficients are given in Table 2.

For Experiment 2, we again analyzed total error numbers. The analysis did not include trial number, nor did it include a random effect of context by items, due to a failure to converge. The model fit was significantly improved by the addition of a fixed effect of context [$\chi^2(3) = 19.6$, $p < .001$]. Inspection of the coefficients revealed fewer errors in the match condition than in the other conditions [$\beta = -0.48$, $SE(\beta) = 0.22$, $z = -2.18$, $p = .029$] and more errors in the mismatch conditions than in the backward control condition [$\beta = 0.96$, $SE(\beta) = 0.27$, $z = 3.59$, $p < .001$]. We found no difference between the rime-overlap and no-overlap conditions ($z = 1.65$). See Table 3 for the model coefficients.

Experiment 3

One participant in Experiment 3 made 22 errors early in the experiment and was excluded from all analyses. The remaining 19 participants made 37 errors over 2,736 responses (1.4 %). Errors and picture-naming latencies are shown in Table 1 and in Fig. 1, respectively. For our analyses, we once again used orthogonal contrasts, this time comparing the matching condition to all others, the two mismatching conditions to the backward control, and finally, the mismatching conditions (onset overlap vs. no overlap) to each other.

Other than the differences in conditions, the models were constructed in the same way as for the two previous experiments. Once again, the fixed effect of context improved the model fit [$\chi^2(3) = 20.4$, $p < .001$]. Participants were quicker to respond as the experiment progressed [$\beta = -15.5$, $SE(\beta) = 5.0$, $t = 3.13$] and responded fastest in the match condition [$\beta = -14.2$, $SE(\beta) = 2.3$, $t = 6.07$]. No other differences were significant (*t*s < 1).

An analysis of total error numbers did not include either trial number or a random effect of context by items. Adding a fixed effect of context significantly improved the model fit [$\chi^2(3) = 12.8$, $p = .005$]. We found no difference between the match condition and the other conditions combined ($z = 1.78$), although the two mismatching conditions resulted in more errors than did the backward control [$\beta = 0.5$, $SE(\beta) = 0.2$, $t = 2.65$]. The onset-overlap and no-overlap conditions did not differ ($z < 1$).

Discussion

In three experiments, we showed that the time to name a picture after hearing a sentence fragment that strongly predicted a given distractor word was faster if the picture name coincided exactly with the distractor than in any other condition. Counter to expectations, we found no response latency differences between the remaining conditions: Naming times did not differ, whether the predicted word overlapped with the picture name at the onset, at the rime, or not at all. Moreover, the naming latencies for these conditions were the same as for a backward control condition in which no specific word could have been predicted from the auditory context. These findings stand in contrast to those from picture–word interference paradigms, in which a distractor is explicitly presented, either in writing (Damian & Dumay, 2007; Lupker, 1982), auditorily (Meyer & Schriefers, 1991), or pictorially (Humphreys et al., 2010). In each of these cases, phonological overlap has been shown to facilitate picture-naming relative to no-overlap conditions. Although the naming latencies reported here were relatively fast, a reliable facilitation effect in the match condition, and the fact that participants' response latencies continued to reduce throughout each experiment, both militate against any suggestion that the lack of phonological effects reported here can be attributed to some kind of ceiling effect. An initial interpretation of the present findings is therefore that, although upcoming material is predicted during comprehension, any involvement of the production system stops short of a speech-sound level of representation.

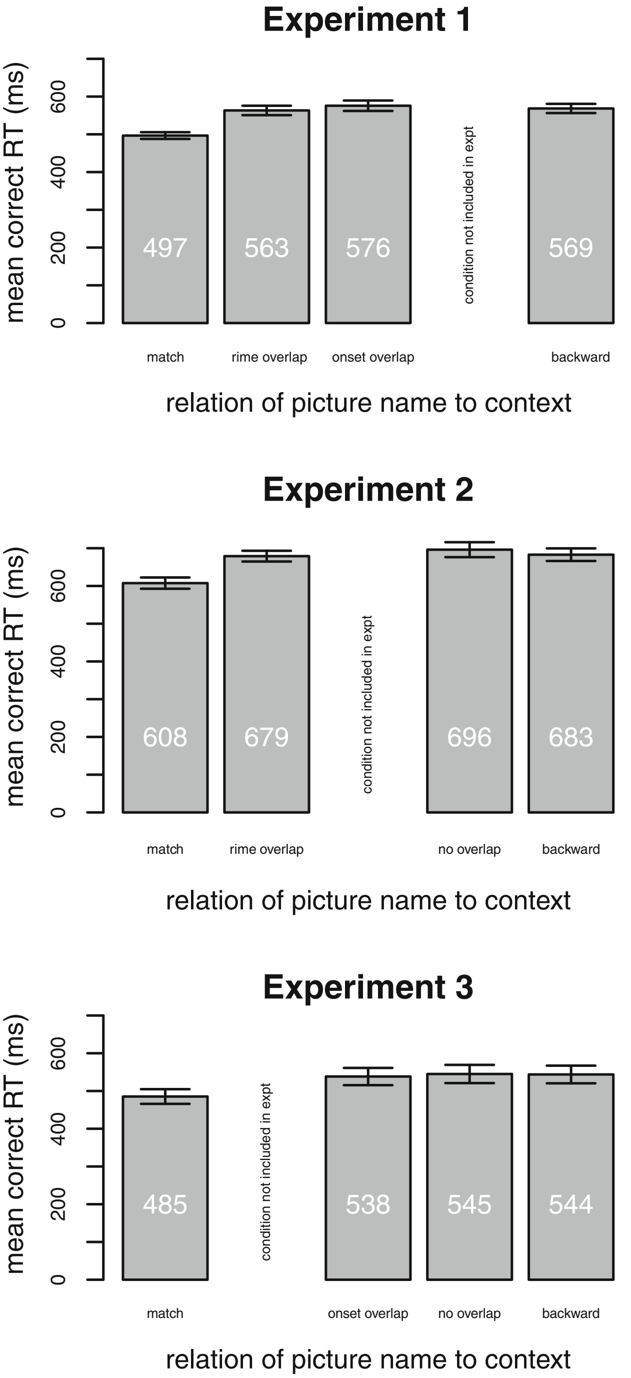

Fig. 1 By-participants mean onset latencies to correctly name pictures in Experiments 1–3. Error bars represent one standard error of the mean.

A potential objection to this interpretation is that the present experiment may not have induced participants to predict specific words at all: Instead, the faster naming of pictures that happened to match predictable words might have been due to the ease with which those picture names could be semantically integrated with the preceding context. In cases of phonological overlap, there would still be a semantic mismatch, and therefore relative difficulty of integration. However, this explanation seems unlikely, because each picture was named several times throughout each of the present experiments. Since the picture names were also used as the words predicted by sentence fragments, it is highly likely that specific words would have been activated at each trial, due to extensive repetition. In line with this suggestion, the majority of recorded errors (60 % and 62 % in Exps. 1 and 2, respectively; 73 % in Exp. 3) were cases in which the words predicted by the sentence fragments were accidentally produced in lieu of the picture names.[2] The fact that distractor names were overtly produced suggests that these names were fully specified at the lexical, and hence the phonological, level. Moreover, the delay in naming pictures in the backward control condition relative to the match condition militates against a purely integration-based account, since in this condition there is no semantic context with which the picture name could be integrated.

Given that distractor words are predicted, the question arises of whether the naming latency differences between the match and other conditions were due to inhibition of mismatching picture names or to facilitation of names that matched. One way of addressing this question would be by considering the backward control conditions in each experiment. In these conditions, the auditory input presented to participants did not allow the generation of predictions at any linguistic level. Thus there was no potential for competitor word representation or competitor-driven interference effects. This is reflected in the fact that error rates in the backward control conditions were significantly lower than those in the mismatch (onset, rime, and no-overlap) conditions, in line with the match conditions. The mismatch conditions, in contrast, clearly elicited potential competitor representations: In these conditions, the name of the competitor was produced in error 103 times across experiments. Despite the differences in error rates, however, response times in the mismatch conditions were equivalent to those in the backward control conditions. In other words, there was evidence that competitors were represented as a result of prediction, but no prima facie evidence that target-naming latencies were affected by inhibitory competition with those competitors.

One way of accounting for this lack of evidence without concluding against inhibition would be to suggest that inhibitory effects were masked by countervailing facilitation. In each of the mismatch conditions, the predicted words had a semantic as well as a phonological mismatch with the target; perhaps phonological facilitation was masked by semantic inhibition? This view is, however, quickly ruled out by the no-overlap conditions in

---

[2] It is worth noting in this context that the prevalence of intrusion errors in the mismatch conditions patterns with picture–word interference studies that have included auditory distractors (Meyer & Schriefers, 1991), but not with studies that have included written distractors (Lupker, 1982, reported fewer than 1 % errors; and Damian & Dumay, 2007, did not observe any differences in numbers of errors across conditions).

Table 2 Model coefficients (in milliseconds) for naming latencies: Experiments 1, 2, and 3

Fixed Effect	Estimate	*SE*	*t*	Random Effect		Variance
Experiment 1						
Intercept	600.53	22.56	26.61	Participants	Intercept	7,367.8
Log(trial pos^n)	–11.74	4.82	–2.44		Log(trial pos^n)	318.4
Context					Context M vs. O	17.2
Match (M) vs. Others (O)	–18.19	2.20	–8.26		Context MM vs. BC	168.4
Mismatch (MM) vs. Backward Control (BC)	–0.28	3.87	–0.07		Context R vs. OO	192.8
Rime (R) vs. Onset Overlap (OO)	–5.55	7.56	–0.74	Items	Intercept	233.7
					Context M vs. O	11.2
					Context MM vs. BC	22.2
					Context R vs. OO	355.7
				Residual		33,039.0
Experiment 2						
Intercept	733.18	16.94	43.29	Participants	Intercept	1,759.2
Log(trial pos^n)	–15.83	4.00	–3.96		Log(trial pos^n)	176.3
Context					Context M vs. O	33.6
M vs. O	–19.66	2.36	–8.33		Context MM vs. BC	88.1
MM vs. BC	1.65	3.71	0.45		Context R vs. NO	300.5
R vs. No Overlap (NO)	–8.67	6.76	–1.28	Items	Intercept	776.2
					Context M vs. O	20.9
					Context MM vs. BC	60.0
					Context R vs. NO	210.4
				Residual		17,808.8
Experiment 3						
Intercept	593.59	29.71	19.98	Participants	Intercept	13,232.5
Log(trial pos^n)	–15.53	4.96	–3.13		Log(trial pos^n)	289.5
Context					Context M vs. O	46.4
M vs. O	–14.15	2.33	–6.07		Context MM vs. BC	182.9
MM vs. BC	–1.03	4.03	–0.26		Context OO vs. NO	31.1
OO vs. NO	–2.44	4.42	–0.55	Items	Intercept	132.2
					Context M vs. O	1.9
					Context MM vs. BC	10.2
					Context OO vs. NO	5.9
				Residual		20,920.8

Experiments 2 and 3. Words predicted in these conditions would have been as vulnerable to semantic inhibition as were words in the other overlap conditions. When vulnerability to inhibition was kept constant in this way, partial phonological overlap between the distractor and the picture name, whether at the onset or the rime, did not affect response times, ruling out any suggestion that inhibition at a semantic level may have been masking phonological-level facilitation effects.[3]

A final possibility is that the present findings should not be considered in terms of activation and inhibition, but instead reflect extralinguistic processes, such as those involved in suppressing a response. Although such a view cannot be completely ruled out, it should be noted that in the backward control conditions there was no competing response to suppress; thus, the arguments that apply to inhibition above also apply to this suggestion.

Taking all of these considerations together, it seems that facilitation, but not inhibition, is implicated in the present experiments. Importantly, it appears that facilitation only occurs when a word that is predicted exactly matches the name to be produced for a picture. If predicted words were represented at a phonological (speech-sound) level within the

---

[3] Note that under a "response exclusion hypothesis" account, under which "semantic" inhibition arises at an articulatory output-buffer level, relative phonological facilitation effects would remain observable (Janssen, 2013; see also Blackford, Holcomb, Grainger, & Kuperberg, 2012, for evidence of phonological facilitation arising from masked primes during picture naming).

Table 3 Model coefficients (in logits) for the likelihood of producing an error: Experiments 1, 2, and 3

Fixed Effect	Estimate	*SE*	*z*	Random Effect		Variance
Experiment 1						
Intercept	–5.43	0.41	–13.3	Participants	Intercept	1.90
Context					Context M vs. O	0.07
Match (M) vs. Others (O)	–0.62	0.15	–4.06		Context MM vs. BC	0.21
Mismatch (MM) vs. Backward Control (BC)	0.87	0.19	4.51		Context R vs. OO	0.15
Rime (R) vs. Onset Overlap (OO)	0.01	0.17	0.07	Items	Intercept	0.26
Experiment 2						
Intercept	–5.28	0.41	–12.82	Participants	Intercept	1.67
Context					Context M vs. O	0.18
M vs. O	–0.48	0.22	–2.18		Context MM vs. BC	0.40
MM vs. BC	0.96	0.27	3.59		Context R vs. NO	0.06
R vs. No Overlap (NO)	0.29	0.17	1.65	Items	Intercept	0.00
Experiment 3						
Intercept	–5.76	0.71	–8.11	Participants	Intercept	0.81
Context					Context M vs. O	0.83
M vs. O	–1.23	0.69	–1.78		Context MM vs. BC	0.06
MM vs. BC	0.47	0.18	2.65		Context OO vs. NO	0.15
OO vs. NO	0.04	0.22	0.16	Items	Intercept	0.00

speech production system, we would expect phonological facilitation to remain apparent; there does not appear to be any evidence that such facilitation was being masked by other processes.

General implications

Over three experiments, we found no evidence to suggest that comprehension-associated predictions gave rise to inhibition within the production system, although we did find evidence of a facilitatory effect. However, this effect does not appear to extend to phonological-level representations accessible to the speech production system: Causing participants to produce words that overlap phonologically with words predictable through comprehension does not give rise to facilitation of production, unless the overlap is complete (and therefore the picture-name matches the predicted word at levels other than phonological). On the assumption that phonology precedes articulation in production, this further implies that the speech–motor system is not involved in making specific predictions.

One way to reconcile these findings with evidence that the speech–motor system is activated during comprehension may be to suggest that this activity is associated with performance rather than with content (e.g., Rothermich & Kotz, 2013; Scott et al., 2009). According to such a view, presentation of an auditory sentence stem would automatically engage prediction processes. One aspect of prediction would be estimation of the timing of the speaker's production, and preparation to respond (related to the so-called "how" pathway), reflected in speech–motor activity. Predictions at the lexical–phonological level (the "what" pathway) would not rely on the speech–motor system (see Hickok, 2012). In the context of the present experiments, the auditory contexts would consistently alert listeners to the likely moment at which a picture might appear. The facilitation effect seen in the full-overlap condition would relate to the "what" channel, and would not involve speech–motor activation.

This interpretation is not inconsistent with DeLong et al.'s 2005) conclusion that phonological-form expectations are generated during the comprehension of high-cloze sentences. It is in no way implicit in that conclusion that such phonological expectations would be accessible to the speech production system. The way that our findings differ from those of DeLong and colleagues recalls the suggestion that phonological representations incorporate separable levels of representation (e.g., Goldrick & Rapp, 2007; Hickok, 2012): Predictions of another's speech may be specified at a phonological level that is driven by lexical or auditory representations, but that is not necessarily implemented at a phonological–articulatory level.

If auditory sentence stems do enable prediction via a "how" pathway, the question remains as to why picture-naming latencies were not relatively delayed in the backward-speech control condition. Backward speech may not offer sufficient cues at either a semantic or a prosodic level from which timing can be estimated (Brown et al., 2012; Londei et al., 2010). Of course, failure to find an effect cannot be interpreted as evidence that there is no such effect; and the possibility remains, in principle at least, that a disadvantage due to backward speech is exactly matched by inhibition in the mismatch conditions. However, in the absence of evidence to support such a view, the present pattern of findings does raise the question of whether any benefits of "how" prediction might be relatively small, at least in the context of the present, rather artificially constrained, task.

Of course, it may be that response latencies do not reliably reflect activation levels through the production system to articulation (although such a suggestion would effectively undermine a substantial body of work). If, however, we align ourselves with the previous literature and accept that the time taken to name a picture is likely to be influenced by preactivation of relevant phonological representations, the present experiments strongly suggest that "what" prediction during comprehension does not appear to occur at a phonological–articulatory level, and thus that the speech–motor activation associated with language comprehension is unlikely to reflect a detailed prediction of the upcoming content.

Author note The research reported in this article was supported in part by an EPSRC Doctoral training studentship awarded to E.D. The authors know of no conflict of interest.

References

Baayen, R. H. (2008). *Analyzing linguistic data (Vol. 505)*. Cambridge, UK: Cambridge University Press.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68,* 255–278. doi:10.1016/j.jml.2012.11.001

Bates, D., Maechler, M., & Bolker, B. (2013). lme4: Linear mixed-effects models using Eigen and S4 [Software] (R package version 0.999999-4). Retrieved from http://cran.r-project.org/package=lme4

Blackford, T., Holcomb, P. J., Grainger, J., & Kuperberg, G. R. (2012). A funny thing happened on the way to articulation: N400 attenuation despite behavioral interference in picture naming. *Cognition, 123,* 84–99.

Brown, E. C., Muzik, O., Rothermel, R., Matsuzaki, N., Juhász, C., Shah, A. K., & Asano, E. (2012). Evaluating reverse speech as a control task with language-related gamma activity on electrocorticography. *NeuroImage, 60,* 2335–2345.

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods, 41,* 977–990. doi:10.3758/BRM.41.4.977

Damian, M. F., & Dumay, N. (2007). Time pressure and phonological advance planning in spoken production. *Journal of Memory and Language, 57,* 195–209.

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience, 8,* 1117–1121. doi:10.1038/nn1504

Fadiga, L., Craighero, L., Buccino, G., & Rizzolatti, G. (2002). Speech listening specifically modulates the excitability of tongue muscles: A TMS study. *European Journal of Neuroscience, 15,* 399–402.

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

Goldrick, M., & Rapp, B. (2007). Lexical and post-lexical phonological representations in spoken production. *Cognition, 102,* 219–260. doi:10.1016/j.cognition.2005.12.010

Greene, J. O. (1988). Cognitive processes: Methods for probing the black box. In C. H. Tardy (Ed.), *A handbook for the study of human communication: Methods and instruments for observing, measuring, and assessing communication processes* (pp. 37–66). Westport, CT: Ablex.

Griffin, Z. M., & Bock, K. (1998). Constraint, word frequency, and the relationship between lexical processing levels in spoken word production. *Journal of Memory and Language, 38,* 313–338.

Hickok, G. (2012). Computational neuroanatomy of speech production. *Nature Reviews Neuroscience, 13,* 135–145.

Hirschfeld, G., Jansma, B., Bölte, J., & Zwitserlood, P. (2008). Interference and facilitation in overt speech production investigated with event-related potentials. *NeuroReport, 19,* 1227–1230. doi:10.1097/WNR.0b013e328309ecd1

Humphreys, K. R., Boyd, C. H., & Watter, S. (2010). Phonological facilitation from pictures in a word association task: Evidence for routine cascaded processing in spoken word production. *Quarterly Journal of Experimental Psychology, 63,* 2289–2296.

Janssen, N. (2013). Response exclusion in word–word tasks: A comment on Roelofs, Piai and Schriefers. *Language and Cognitive Processes, 28,* 672–678.

Jescheniak, J. D., Oppermann, F., Hantsch, A., Wagner, V., Mädebach, A., & Schriefers, H. (2009). Do perceived context pictures automatically activate their phonological code? *Experimental Psychology, 56,* 56–65. doi:10.1027/1618-3169.56.1.56

Londei, A., D'Ausilio, A., Basso, D., Sestieri, C., Gratta, C. D., Romani, G. L., & Belardinelli, M. O. (2010). Sensory–motor brain network connectivity for speech comprehension. *Human Brain Mapping, 31,* 567–580.

Lupker, S. J. (1982). The role of phonetic and orthographic similarity in picture–word interference. *Canadian Journal of Psychology, 36,* 349–367. doi:10.1037/h0080652

Mahon, B. Z., Costa, A., Peterson, R., Vargas, K. A., & Caramazza, A. (2007). Lexical selection is not by competition: A reinterpretation of semantic interference and facilitation effects in the picture–word interference paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33,* 503–535. doi:10.1037/0278-7393.33.3.503

Meyer, A. S., & Damian, M. F. (2007). Activation of distractor names in the picture-picture interference paradigm. *Memory & Cognition, 35,* 494–503. doi:10.3758/BF03193289

Meyer, A. S., & Schriefers, H. (1991). Phonological facilitation in picture–word interference experiments: Effects of stimulus onset asynchrony and types of interfering stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 1146–1160. doi:10.1037/0278-7393.17.6.1146

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. doi:10.1037/0278-7393.28.3.555

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.

Nozari, N., Dell, G. S., & Schwartz, M. F. (2011). Is comprehension necessary for error detection? A conflict-based account of monitoring in speech production. *Cognitive Psychology, 63,* 1–33. doi:10.1016/j.cogpsych.2011.05.001

Oppermann, F., Jescheniak, J. D., & Görges, F. (2014). Resolving competition when naming an object in a multiple-object display. *Psychonomic Bulletin & Review, 21,* 78–84. doi:10.3758/s13423-013-0465-5

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences, 27,* 169–190. disc. 190–226. doi:10.1017/S0140525X04000056

Pickering, M. J., & Garrod, S. (2007). Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences, 11,* 105–110.

Pulvermüller, F., Huss, M., Kherif, F., del Prado Martin, F. M., Hauk, O., & Shtyrov, Y. (2006). Motor cortex maps articulatory features of speech sounds. *Proceedings of the National Academy of Sciences, 103*, 7865–7870.

R Development Core Team. (2014). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. Retrieved from www.r-project.org

Raney, G. E. (1993). Monitoring changes in cognitive load during reading: An event-related brain potential and reaction time analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 51–69. doi:10.1037/0278-7393.19.1.51

Roe, K., Jahn-Samilo, J., Juarez, L., Mickel, N., Royer, I., & Bates, E. (2000). Contextual effects on word production: A lifespan study. *Memory & Cognition, 28,* 756–765.

Rothermich, K., & Kotz, S. A. (2013). Predictions in speech comprehension: fMRI evidence on the meter–semantic interface. *NeuroImage, 70,* 89–100. doi:10.1016/j.neuroimage.2012.12.013

Schiller, N. O., Horemans, I., Ganushchak, L., & Koester, D. (2009). Event-related brain potentials during the monitoring of speech errors. *NeuroImage, 44,* 520–530. doi:10.1016/j.neuroimage.2008.09.019

Scott, S. K., McGettigan, C., & Eisner, F. (2009). A little more conversation, a little less action—Candidate roles for the motor cortex in speech perception. *Nature Reviews Neuroscience, 10,* 295–302.

Severens, E., Janssens, I., Kühn, S., Brass, M., & Hartsuiker, R. J. (2011). When the brain tames the tongue: Covert editing of inappropriate language. *Psychophysiology, 48,* 1252–1257.

Watkins, K., & Paus, T. (2004). Modulation of motor excitability during speech perception: The role of Broca's area. *Journal of Cognitive Neuroscience, 16,* 978–987.

Watkins, K. E., Strafella, A. P., & Paus, T. (2003). Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia, 41,* 989–994.

Wicha, N. Y. Y., Orozco-Figueroa, A., Reyes, I., Hernandez, A., de Barreto, L. G., & Bates, E. A. (2005). When zebras become painted donkeys: Grammatical gender and semantic priming interact during picture integration in a spoken Spanish sentence. *Language and Cognitive Processes, 20,* 553–587. doi:10.1080/01690960444000241