Published in final edited form as:

Lang Cogn Neurosci. 2020 ; 35(8): 980–991. doi:10.1080/23273798.2019.1686529.

Investigating the effects of phonological neighbors on word retrieval and phonetic variation in word naming and picture naming paradigms

Haoyun Zhang, Matthew T. Carlson, Michele T. Diaz

Pennsylvania State University

Abstract

Phonological neighbors have been shown to affect word processing. Prior work has shown that when a word with an initial voiceless stop has a contrasting initial voiced stop neighbor, Voice Onset Times (VOTs) are longer. Higher phonological neighborhood density (PND) has also been shown to facilitate word retrieval latency, and be associated with longer VOTs. However, these effects have rarely been investigated with picture naming, which is thought to be a more semantically driven task. The current study examined the effects of phonological neighbors on word retrieval times and phonetic variation, and how these effects differed in word naming and picture naming paradigms. Results showed that PND was positively correlated with longer VOT in both paradigms. Furthermore, the effect of initial stop neighbors on VOTs was only significant in word naming. These results highlight the influence of phonological neighbors on word production in different paradigms, support interactive models of word production, and suggest that hyper-articulation in speech does not solely depend on communicative context.

Keywords

Language production; Interactive effects; Phonological neighborhood density; Minimal Pair; Voice Onset Time

Introduction

Speaking, or language production, is a fundamental aspect of communication that involves several processes: activating semantic information, selecting the correct lexical entry from the mental lexicon, retrieving phonological information, phonetic encoding, and articulation (Burke & Shafto, 2008; Dell & O'Seaghdha, 1992; Levelt, 1999; Levelt, Roelofs, & Meyer, 1999; Martin, 2003; Schwartz, Dell, Martin, Gahl, & Sobel, 2006). Although the above-mentioned processes are distinct, many word production models suggest that these stages are highly interactive (e.g., Dell, 1986; Dell, Schwartz, Martin, Saffran, & Gagnon, 1997; Goldrick, 2006; Rapp & Goldrick, 2000).

Correspondence concerning this article should be addressed to: Michele T. Diaz, Department of Psychology, 356 Moore Building, University Park, PA, USA, 16801. Tel: +1 814 863 1726. Fax: +1 814 863 7002.

Data and analysis scripts are available through the Open Science Framework: https://osf.io/2cdjz/?view_only=219ae7e45d314b6da7c70b3384cb22db

One of the most well-established models of language production is Dell and colleagues' (1997) two-step interactive activation model. The first step is lemma access, which involves both semantic processing and mapping concepts to the mental lexicon (also referred to as lexical processing). The second step is phonological processing, which involves retrieving the phonological frame of a word and articulation (also referred to as postlexical processing). Interactive models suggest that these processes are interactive where the activation of any one process can spread to and influence the activation of other processes in turn. On the other hand, feed-forward models of language production (Levelt, 1999) consist of similar processes, but activation only flows from early to later processes. In other words, feed-forward models argue that activation of phonological information cannot spread back to the activation of word forms, which cannot spread back to lemma level activation.

Abundant research has provided evidence for models of word production, by investigating the effects of different word characteristics on word retrieval. For instance, studies have shown that semantic variables (e.g., imageability) affect word naming speed, suggesting feed-back activation from word forms to conceptual information, then back to lexical processing (Shibahara et al., 2003; Strain, Patterson, & Seidenberg, 1995). Likewise, lexical characteristics such as word frequency and naming agreement can also affect word retrieval times (e.g., Barry, Morrison, & Ellis, 1997; Carroll & White, 1973).

Among various word characteristics that modulate word retrieval, the current study focuses on the effects of phonological neighbors. Phonological neighbors are words that can be formed from a given word by substituting, adding, or deleting one phoneme. Phonological aspects of production are of interest as these processes undergo age-related decline (Burke, MacKay, Worthley, & Wade, 1991; Burke & Shafto, 2008; Diaz, Johnson, Burke, & Madden, 2014; Rizio, Moyer, & Diaz, 2017). Moreover, in younger adults, phonological neighborhood density (PND; i.e., the number of phonological neighbors) has been shown to significantly affect word retrieval latency and accuracy in most word naming and some picture naming paradigms, displaying either inhibitory effects (Sadat, Martin, Costa, & Alario, 2014) or more often facilitation effects (Adelman & Brown, 2007; Baus, Costa, & Carreiras, 2008; Mirman, Kittredge, & Dell, 2010; Vitevitch, 2002), which might be subject to the particularities of word formation in specific languages (Vitevitch & Stamer, 2006). The effect of phonological neighbors on word retrieval supports interactive models of language production. Specifically, in interactive models, the activation of the target word's phonological units spreads to phonological neighbors of the target word, which in turn spreads among neighbors and back to the target word's phonological units. Because these phonological neighbors are similar to the target word's phonological representations, target word retrieval will be affected by the activation of its phonological neighbors. These effects cannot be accounted for by feed-forward models of language production, as they do not allow any backward influence from phonological segments to word forms. Additionally, other research has shown that higher phonological neighborhood density produces lexically conditioned phonetic variation such as longer voice onset times (VOTs, i.e., the length of time that passes between the release of a stop consonant and the onset of voicing; Fox, Reilly, & Blumstein, 2015), more coarticulation (Scarborough, 2013; Scarborough & Zellou, 2012) and more expanded vowel spaces (Munson & Solomon, 2004; Wright, 2004), which has been suggested to reflect production-internal interactions (i.e., the structure of

interactions among processes within the production system, Baese-Berk & Goldrick, 2009) or increased contextual confusability (Buz, Tanenhaus, & Jaeger, 2016).

Although phonological neighbors are generally considered to be words differing from each other by one phoneme (addition, deletion, or substitution), the difference can be as small as a single phonetic unit, such as the voicing of the initial consonant (e.g., *cape* – *gape*, which begin with voiceless and voiced velar stops, respectively). We will distinguish between such close minimal pairs (henceforth “minimal pairs”), and phonological neighbors more generally, because the existence of a close minimal pair has been linked to phonetic variation in naming words. For instance, two recent studies (Baese-Berk & Goldrick, 2009; Peramunage, Blumstein, Myers, Goldrick, & Baese-Berk, 2011) asked participants to overtly read words with initial voiceless stop consonants to investigate how the presence of a phonetic minimal pair neighbor with a contrasting initial voiced stop affects voice onset time. These two studies reported that the VOTs of words with initial voiceless stop consonants were longer in words that had a contrasting initial voiced stop neighbor than words that did not have such a neighbor (e.g., *cake* does not have a neighbor **gake*). It was suggested that this effect may arise from spreading activation from a close voiced stop neighbor which affected the articulation of the target word that had an initial voiceless stop. Furthermore, Fricke and colleagues (2016) re-analyzed the dataset from Baese-Berk and Goldrick (2009) to investigate the effect of phonological neighbors on the VOTs of minimal pair and non-minimal pair words. They found that both the location of the overlap between neighbors and target words and the total number of phonological neighbors contributed significantly to the VOTs of the target words.

Although there is considerable evidence supporting the influence of phonological attributes on phonetic variation, there has been debate about the underlying mechanisms. For example, a number of studies have suggested that the hyper-articulation effect (e.g., increased VOTs for voiceless stops) that occurs when a close competitor exists may also be a function of communication context (Buz et al., 2016; Scarborough & Zellou, 2013). Specifically, speakers might produce hyper-speech when factors in a communicative environment place extra demands on listeners. For instance, researchers have found that when listeners misunderstood speech, the size of the hyper-articulation effect significantly increased when a phonetic competitor was presented (Buz et al., 2016; Schertz, 2013). These results suggest that the hyper-articulation effect may serve as a way to clarify speech for the listeners’ benefit. However, studies focusing on natural speech also showed that the existence of a voiced-stop minimal pair predicted significantly longer VOTs, even when no listener was involved (Nelson & Wedel, 2017; Wedel, Nelson, & Sharp, 2018). It may be the case that long-term exposure to hyper-articulated VOTs from speech with listeners could lead to differences in the target pronunciations of those words, even when there is no listener present. Therefore, it is still unclear if the hyper-articulation effect in speech is for the listeners’ benefits or just a by-product of speech.

Although there is debate about the nature of these effects, most of the evidence reviewed above supports interactive models of language production through the effect of phonological neighbors on word retrieval times and lexically conditioned phonetic variation. This is because, in strictly feed-forward models, the activation of phonology proceeds automatically

after a word's lexical information is selected. Therefore, phonological neighbors of a word cannot be activated or further affect word production. On the other hand, interactive models of language production allow the activation of phonological segments of the target word to feed back to activate other words who share these phonological segments, further affecting the production of the target word.

When exploring the effect of phonological attributes on word retrieval, most previous studies have used either word naming or picture naming paradigms. While both paradigms examine word production, the influence of various processes differs across the paradigms. Specifically, picture naming involves a much higher extent of feed-forward activation from the semantic level to the lexical level compared to word naming. On the other hand, word naming is a more orthographically driven paradigm compared to picture naming as the word form is provided in word naming. In other words, word naming explicitly provides the orthographic information, providing a route to phonology without necessarily activating semantics. Therefore, a direct comparison between the two paradigms on the effects of phonological neighbors on word retrieval would inform both task driven influences on phonological processes and theoretical accounts of language production.

In the current study, we systematically examined the effects of phonological neighborhood density and minimal pair status on word retrieval times (i.e., reaction times) and phonetic variations (i.e., VOTs), and how these effects differed in a picture naming paradigm (Experiment 1) and a word naming paradigm (Experiment 2). Moreover, we controlled for several lexical and phonetic characteristics (including word frequency, number of syllables, name agreement in picture naming, average biphone probability, and first vowel height). We hypothesized that phonological neighborhood density and minimal pair status would affect both word retrieval times and phonetic variation, which would support interactive accounts of language production. Additionally, the effect of minimal pair status on word production should be stronger in word naming compared to picture naming, considering that picture naming is a more semantically driven task. In particular, although both picture and word naming involve similar processing steps (i.e., semantic activation, lexical retrieval, phonological encoding), the relative emphasis on each process varies across paradigms. In the case of a semantically driven process, such as picture naming, the effect of feed-forward activation from semantics to lexical selection would be much stronger than it would be in word naming, where a direct orthography-phonology route is available. Additionally, in the case of word naming, where the word form is presented, the activation of a contrasting neighbor with a very similar form and its feed-back activation should be very strong.

Finally, to help understand the different processes involved in picture naming and word naming, and to clarify different models of language production, a direct comparison between the two paradigms would also speak to the relationship between hyper-articulation and communication contexts. If hyper-articulation occurs for the purpose of clarifying speech for the listeners' benefit, we should not see any difference between the two paradigms given that the communication contexts of the two paradigms were the same (i.e., no listener present or feedback provided). On the other hand, if the relationship between phonological neighbors and VOT differs between picture naming and word naming, it would indicate that hyper-articulation in speech does not depend solely on communication contexts.

Experiment 1: Picture Naming

Methods

Participants—Fifty college students participated in this experiment. One was excluded from the analysis because the microphone did not pick up most of the responses due to a soft voice, leaving 49 data sets for subsequent analyses. All participants had normal or corrected-to-normal vision and reported no psychiatric or neurological illnesses. They were all native American English speakers with little knowledge of other languages. All participants gave written, informed consent, and all procedures were approved by the Institutional Review Board at the Pennsylvania State University.

Stimuli and Procedure—Participants completed a picture naming task. Photographs were presented and participants were instructed to overtly name the photograph as quickly and accurately as possible. Target names of photographs began with a voiceless stop consonant. Because VOTs needed to be measured, only target words starting with /p/, /t/, and /k/ were used as critical stimuli. There were two conditions: minimal pair (MP) and non-minimal pair (Non-MP). The MP condition consisted of pictures with target names with voiceless initial stops that have a neighbor with a voiced initial consonant (e.g., target word *cape* has a voiced neighbor *gape*). The Non-MP condition was created by pairing every MP word with a non-minimal pair word that has the same stop consonant and a similar first vowel¹, which lacked such a neighbor (e.g., target word *cake* does not have a voiced neighbor **gake*). There were 24 items in each condition and all words started with a CVC format (Consonant-Vowel-Consonant, e.g., *cape* vs. *cake*). Thirty filler pictures whose primary names started with other consonants were also included to obscure the experimental hypotheses and to provide a richer phonological set of picture names for participants to produce. For each trial, a fixation cross first appeared on a white background for 1000 ms, followed by a color photograph of an object or action. Participants were instructed to respond with the photograph's name, using either a noun or a verb. The photograph disappeared immediately after participants made a response or when the maximum response time of 3000 ms was reached. This was followed by a blank screen (duration = 1000 ms). Before the critical trials, participants underwent a practice run consisting of 10 pictures. Stimuli were not repeated across the practice run or experimental conditions. Participants' reaction times were measured and their responses were recorded using a microphone and a digital recorder.

Photographs were taken from normed databases (Brodeur, Guérard, & Bouras, 2014; Moreno-Martínez & Montoro, 2012) and online resources, and depicted a broad range of common objects and actions. Additionally, we normed the photographs with an initial set of 71 MP and Non-MP words with an independent group of 21 healthy, native American English-speaking adults. We then selected 24 pairs (48 words) which had naming consistencies of 61% or higher. The linguistic characteristics (e.g., word frequency, number of syllables, heights of the first vowel, phonological neighborhood density) of the photograph names were obtained from the International Phonetic Alphabet (IPA) chart and

¹For most word pairs but not all of them, the first vowel is the same. Therefore, the vowel height was included as a control variable when analyzing VOTs.

English Lexicon Project (ELP, Balota et al., 2007). The average biphone probability was obtained using the Phonotactic Probability Calculator (Vitevitch & Luce, 2004). For each item, an H-index ($-\sum_{i=1}^k p_i \log_2(1/p_i)$, where k is the number of different names produced to a picture, and p_i is the proportion of participants producing the i th name), a measure of naming consistency or agreement (Snodgrass & Vanderwart, 1980), was calculated based on the responses from the 49 participants who participated in Experiment 1.

Data Analyses—Stimuli in the two critical minimal pair conditions (MP and Non-MP) were included in the analysis (i.e., 24 words in each condition). Item-level H-index was calculated based on the number of acceptable alternatives for each item and the proportion of participants who produced each alternative. An H-index of 0 reflects perfect name agreement and larger H-index indicates lower name agreement (Snodgrass & Vanderwart, 1980). Response accuracy was coded based on the recordings from the session. Responses were marked as correct only if the participant provided the exact target name (e.g., *cap* for *cap*) or plural forms of the same word (e.g., *pears* for *pear*). Other responses, hesitations, or omissions were coded as incorrect and comprised 14.13% of trials. Due to this very strict criterion, all items had an accuracy higher than 40% (Two words' accuracy was lower than 50%). Only correct trials were included in the analyses of reaction time (RT) and voice onset time (VOT).

Prior to analyses, RTs were trimmed – any RTs longer or shorter than 2.5 standard deviations from the individual's overall mean or shorter than 200 ms were excluded (2.49 % of trials were thus considered outliers and excluded). For each MP and Non-MP stimulus, the VOTs of the initial voiceless stop consonant (i.e., /p/, /t/, /k/) were coded by four independent coders using PRAAT (Boersma & Weenink, 2002). The VOT of a word was calculated as the duration from the onset of the burst to the onset of the first vowel². To ensure reliability in data coding, 10% of the data across the two experiments was randomly selected and coded by all four coders. The inter-coder agreement of VOTs reached a very high level (ICC = .96; Based on Koo & Li, 2016, ICC values greater than 0.9 indicate excellent reliability).

RTs, VOTs, and accuracies were analyzed with generalized linear mixed-effect modeling, employing *lmer* and *glmer* functions in the *lme4* package, respectively (Bates, Mächler, Bolker, & Walker, 2014) in the R environment (R Core Team, 2014). Unlike ANOVAs, this approach has the advantage of considering individual data points and controls for variation across participants and items simultaneously, producing more generalizable results. For each dependent variable, we began with a basic model that included fixed slopes of control variables (i.e., H-index, word frequency, number of syllables, and average biphone probability in all models, and reaction time³ and first vowel heights in VOT models), random intercepts by participant and by word, and random slopes (by participant) of

²VOTs were measured from the point when a vertical striation in the spectrogram and amplitude spike in the waveform were evident to the point when the waveform became consistently periodic and the spectrogram showed clear format structure. When there was a double burst, the second one was marked. Breathiness was less of our concern because we were dealing with single word in English – word initial voiceless aspirated stops followed by vowels. A detailed coding manual is available through the Open Science Framework (https://osf.io/2cdjz/?view_only=219ae7e45d314b6da7c70b3384cb22db).

³To make sure that the VOT effect is not just a by-product of reaction time, we added reaction time as one control variable in VOT regression models. In fact, VOT and reaction time only significantly correlated with each other in Experiment 2 (Word Naming, p

phonological neighborhood density and minimal pair condition (MP vs. Non-MP)⁴. Next, we followed a stepwise procedure, adding the fixed effect of either phonological neighborhood density or minimal pair condition, and then the other of these two variables. The analysis was performed using both stepwise orders because Condition (MP vs. Non-MP) and PND were related: words in the MP condition had significantly greater phonological neighborhood density than words in the Non-MP condition ($p < .001$). This analysis allowed us to see whether either variable accounted for additional variance, above that shared by both. We used the ANOVA function to compare models and decide whether the added independent variable significantly improved the model log-likelihood or not (Barr et al., 2013; R Core Team, 2014). In terms of variable distribution, a general rule of thumb is that the data is considered as fairly symmetrically distributed if the skewness is between -0.5 and 0.5 . Because the distribution of RTs was very skewed (Supplemental Figure 1a; skewness = 1.50), they were log-transformed (skewness = 0.57 after transformation). VOTs were not transformed because their distribution was not skewed (Supplemental Figure 1c; skewness = 0.24). Minimal pair condition (MP vs. Non-MP), and first vowel height (low vs. high, with no mid vowels) were contrast coded (-0.5 vs. 0.5). Continuous variables included H-index, number of syllables of the target word, target word log frequency, and target word phonological neighborhood density. Continuous variables were z-scored.

Results

Four figures were plotted to demonstrate the effects of the two critical phonological variables on both reaction time and voice onset time (See Figure 1 for the effect of PND on RT, Figure 2 for the effect of MP on RT, Figure 3 for the effect of PND on VOT, and Figure 4 for the effect of MP on VOT). To facilitate comparison, each plot included the results of both Experiment 1 (Panel a) and Experiment 2 (Panel b). Values shown in the figures were observed values of dependent variables. In short, we found that higher phonological neighborhood density was associated with longer VOTs in both experiments, and MP words had longer VOTs than Non-MP words in word naming.

Reaction Times—For the picture naming task, the basic model of reaction time included fixed slopes of control variables (i.e., H-index, word frequency, number of syllables, and average biphone probability), random intercepts by participant and by word, and random slopes of phonological neighborhood density and minimal pair condition (MP vs. Non-MP). The final fitted basic models can be found in Supplemental Table 1A. The model was not significantly improved either by adding phonological neighborhood density to the basic model ($\chi^2 = .05$, $df = 1$, $p = .82$), or by adding minimal pair condition in addition to phonological neighborhood density ($\chi^2 = 1.42$, $df = 1$, $p = .23$). In addition, adding minimal pair condition to the basic model did not significantly improve the model fit ($\chi^2 = 1.50$, $df = 1$, $p = .22$), and adding phonological neighborhood density in addition to minimal pair

$< .001$), but not in Experiment 2 (Picture Naming, $p = .70$), which suggests that the relationship between these two factors was driven by different language processes. More details can be found in the discussion. Additionally, we also ran all the VOT models without RT and included these in the Supplemental Materials (See Supplemental Materials section *VOT effects without including RT as a covariate* for details). These results were largely consistent with the models while including RT.

⁴Note that for cases where full models did not converge, we took out subject level random intercept but keeping the random slopes (Barr, Levy, Scheepers, & Tily, 2013). For cases like this, we also took out subject level random intercept of the model in comparison to make sure the two models are comparable in terms of model fits.

condition did not significantly improve the model fit either ($\chi^2 = .08$, $df = 1$, $p = .78$). In summary, neither phonological neighborhood density (Figure 1a) nor minimal pair condition (Figure 2a) significantly predicted reaction times in the picture naming task.

Accuracy—The mean accuracy across all items and participants was 88.01%. A mixed logistic regression was conducted on the number of response errors to explore the effect of phonological neighborhood density and minimal pair condition. A basic model of accuracy included the same variables as the reaction time model (See Supplemental Table 1A for full fitted model details). Adding phonological neighborhood density ($\chi^2 = .25$, $df = 1$, $p = .62$) or the minimal pair condition ($\chi^2 = 1.12$, $df = 1$, $p = .29$) to the basic model did not significantly improve the model fit. Adding the one variable in addition to the other, did not improve the model fit either (Adding MP condition to PND: $\chi^2 = 1.65$, $df = 1$, $p = .20$; Adding PND to MP condition: $\chi^2 = .78$, $df = 1$, $p = .38$). In summary, similar to reaction time models, neither phonological neighborhood density nor minimal pair condition significantly predicted picture naming accuracy.

VOT—A linear basic mixed-effect model on VOTs included fixed slopes of control variables (i.e., H-index, word frequency, number of syllables, average biphone probability, first vowel height, and log-transformed reaction time), random intercepts by participant and by word, and random slopes of phonological neighborhood density and minimal pair condition (MP vs. Non-MP). The final fitted basic model can be found in Supplemental Table 1A. The log-transformed RT was included in the model to account for the potential carry-over effect of word retrieval on VOTs⁵. Adding phonological neighborhood density to the basic model significantly improved the model fit ($\chi^2 = 6.55$, $df = 1$, $p = .01$). This result indicated that phonological neighborhood density was a significant predictor of VOTs in picture naming. Adding minimal pair condition in addition to PND did not significantly improve the model fit ($\chi^2 = .001$, $df = 1$, $p = .97$). On the other hand, adding minimal pair condition to the basic model did not significantly improve the model fit ($\chi^2 = 1.17$, $df = 1$, $p = .28$), but adding phonological neighborhood density in addition to MP condition consistently improved the model fit ($\chi^2 = 5.38$, $df = 1$, $p = .02$). In summary, higher phonological neighborhood density was associated with longer VOTs (Figure 3a), while minimal pair condition did not significantly predict the VOTs in picture naming (Figure 4a).

Experiment 2: Word Naming

Methods

Participants—A different group of 51 college students with comparable characteristics to those in Experiment 1, participated in this experiment. All participants gave written, informed consent, and all procedures were approved by the Institutional Review Board at the Pennsylvania State University.

⁵Additionally, the reaction times in picture naming were overall much slower than the reaction times in word naming. Adding reaction time in VOT models would ultimately help control for some extraneous task demands that may affect reaction times, especially when comparing the two tasks in later analyses.

Stimuli and Procedure—The same set of stimuli was used in this experiment (24 MP vs. Non-MP pairs, 30 filler words, and 10 practice words). However, in this experiment, instead of naming pictures, participants were presented with words and asked to read each word aloud as it appeared on the screen. Each trial started with a black fixation cross (duration = 1000 ms), followed by the presentation of a word (presented in black, 36pt, Courier New, font on a white background). The word disappeared after the participant made a response or after 2000 ms had elapsed. This was followed by a blank screen (duration = 1000 ms).

Data Analyses—RTs, VOTs, and accuracies were coded using the same criteria as Experiment 1. 0.42 % of the trials were excluded because of incorrect responses and 2.61 % of RTs were excluded after trimming.

Multi-level modeling analyses similar to those from Experiment 1 were conducted on RTs and VOTs. RTs were log-transformed because they were positively skewed (Supplemental Figure 1b; skewness = 0.79; after transformation skewness = 0.07). VOTs were not transformed because their distribution was not skewed (Supplemental Figure 1d; skewness = 0.45). However, no further analyses were conducted on accuracies because participants' performance was at ceiling (i.e., 99.58% of responses were correct, the lowest item level accuracy was 96.08%). Categorical variables were contrast coded and continuous variables were z-scored, as in Experiment 1. H-index was not included in this analysis as the task was word reading, and there was virtually no naming variability.

Results

Reaction Times—A linear mixed-effects model was conducted on word naming RTs to explore the effect of phonological neighborhood density and minimal pair condition. The basic model included fixed effects of word frequency, number of syllables, and average biphone probability, random intercepts of subject and word, and random slopes (by subject) of phonological neighborhood density and minimal pair condition. The final fitted basic models can be found in Supplemental Table 1B. Neither phonological neighborhood density (Figure 1b, $\chi^2 = .66$, $df = 1$, $p = .42$) nor minimal pair condition (Figure 2b, $\chi^2 = .22$, $df = 1$, $p = .64$) significantly improved the model fit over the baseline model. Moreover, adding one factor in addition to the other did not significantly improve the model fit either (Adding PND in addition to MP condition: $\chi^2 = 1.19$, $df = 1$, $p = .28$; Adding MP condition in addition to PND: $\chi^2 = .75$, $df = 1$, $p = .39$).

VOTs—A linear mixed-effect model was conducted on VOTs. In addition to the variables included in reaction time models, the basic model of VOT also included fixed effects of first vowel height, and log-transformed reaction time. The final fitted basic models can be found in Supplemental Table 1B. Results showed that adding phonological neighborhood density to the basic model significantly improved the model fit ($\chi^2 = 14.33$, $df = 1$, $p < .001$), then adding minimal pair condition in addition to PND did not improve the model fit further ($\chi^2 = .75$, $df = 1$, $p = .39$). On the other hand, adding minimal pair condition to the basic model first significantly improved the model fit ($\chi^2 = 4.13$, $df = 1$, $p = .04$), then adding phonological neighborhood density in addition to MP condition further improved the model fit ($\chi^2 = 10.95$, $df = 1$, $p < .001$). These results suggested that higher phonological

neighborhood density was associated with longer VOTs in word naming (Figure 3b). Additionally, MP words had significantly longer VOTs compared to Non-MP words (Figure 4b), although this effect might be accounted by its shared variance with the effect of PND on VOTs.

Additional Analyses Comparing the Two Paradigms—In order to compare the results of the two experiments directly, we built additional models of RT and VOT using the combined results from both picture naming and word naming experiments.

Reaction Times—The basic model of reaction time included H-index and its interaction with paradigm⁶, word frequency, number of syllables, and average biphone probability, random intercepts by subject and word, and random slopes (by subject) of phonological neighborhood density and minimal pair condition. The final fitted basic models can be found in Supplemental Table 1C. Using a stepwise regression, phonological neighborhood density (Figure 1; $\chi^2 = .62$, $df = 1$, $p = .43$), the interaction between PND and paradigm ($\chi^2 = 2.19$, $df = 1$, $p = .14$), minimal pair condition ($\chi^2 = .78$, $df = 1$, $p = .38$), and the interaction between MP condition and paradigm ($\chi^2 = 2.17$, $df = 1$, $p = .14$) were added to the model in that order. None of the factors significantly improved the model fit. Alternatively, the main effect of MP condition and its interaction with the paradigm were added to the basic model first, then the effect of PND and its interaction with the paradigm were added on top. Results showed that only the interaction effect between MP condition and paradigm was a significant predictor of reaction times before PND effects were added to the model (MP condition: $\chi^2 = .24$, $df = 1$, $p = .62$; interaction between MP condition and paradigm: $\chi^2 = 4.07$, $df = 1$, $p = .04$; PND: $\chi^2 = 1.35$, $df = 1$, $p = .25$; and interaction between PND and paradigm: $\chi^2 = .10$, $df = 1$, $p = .76$). Further analyses were conducted to explore this interaction. In terms of paradigm difference in each word type, the reaction times in picture naming were significantly longer than the word naming in both MP words and Non-MP words ($ps < .001$). On the other hand, although the effect of MP condition on reaction time was not significant in either paradigm, there was a trend in the picture naming paradigm as indicated in Figure 2a.

VOTs—The basic model of VOT included fixed effects of H-index and its interaction with paradigm, word frequency, number of syllables, average biphone probability, first vowel height, and log-transformed reaction time, random intercepts of word and subject, and random slopes (by subject) of phonological neighborhood density and minimal pair condition. The final fitted basic models can be found in Supplemental Table 1C. First, factors including PND ($\chi^2 = 10.47$, $df = 1$, $p = .001$), the interaction between PND and paradigm ($\chi^2 = 3.30$, $df = 1$, $p = .07$), MP condition ($\chi^2 = .15$, $df = 1$, $p = .70$), and the interaction between MP condition and paradigm ($\chi^2 = 2.10$, $df = 1$, $p = .15$) were added to the model in that order. Only the effect of PND on VOTs was significant, consistent with the results across both paradigms, indicating that higher phonological neighborhood density was associated with longer VOTs (Figure 3). Additionally, these variables were added to the basic model stepwise in a different order (MP condition: $\chi^2 = 2.19$, $df = 1$, $p = .14$;

⁶We added the interaction between H-index and paradigm because H-index was only expected to affect word retrieval in picture naming (i.e., there was almost no variability in word naming responses).

interaction between MP condition and paradigm: $\chi^2 = 3.81$, $df = 1$, $p = .051$; PND: $\chi^2 = 8.60$, $df = 1$, $p = .003$; interaction between PND and paradigm: $\chi^2 = 1.38$, $df = 1$, $p = .24$). In addition to the significant effect of PND on VOTs, the interaction effect between MP condition and paradigm on VOTs was nearly significant. This is consistent with results from the individual experiments in which the effect of MP condition on VOTs (i.e., MP words had longer VOTs compared to Non-MP words) was only significant in word naming, but not in picture naming (Figure 4). However, the effect of ordering may reflect shared variance between MP condition and phonological neighborhood density, with the effects of phonological neighborhood density accounting for unique variance beyond that shared with MP, as PND was significant in either ordering.

Discussion

The primary goal of the current project was to investigate the effects of phonological factors, such as phonological neighborhood density and minimal pair status, on word retrieval and phonetic variation, and how these effects were modulated by different paradigms such as picture naming and word naming. In general, our results support interactive accounts of word production. They also suggest that the hyper-articulation effect in speech does not solely depend on speech context and may be task dependent.

First, no significant effect was found for phonological neighborhood density on naming latencies in either picture naming or word naming. Some previous studies have reported that word naming times are faster for words from dense neighborhoods compared to words from sparse neighborhoods, reflecting a facilitation effect on word retrieval from phonological neighbors (Andrews, 1989, 1992; Baus et al., 2008; Mirman et al., 2010; Vitevitch, 2002). However, other studies have reported an interference effect from phonological neighbors on naming latencies and argued that aspects of phonological neighbors (e.g., neighborhood frequency, onset density, etc.) mediate the effect of phonological neighborhood density on naming latencies (Sadat et al., 2014). Given the literature, it is surprising, if not unprecedented, that the current study did not find any significant effect of phonological neighborhood density on word retrieval times. The lack of an effect in picture naming might be because the effect of phonological neighborhood density on reaction time interacted with the minimal pair condition. For instance, words with an onset neighbor, as in the MP condition, may be less sensitive to the presence of other neighbors due to stronger interference from the onset neighbor, either because it occurs directly at the word's onset, or because the words we have identified as onset neighbors in this condition differed by only a single phonological feature (place of articulation), whereas other neighbors might involve greater contrasts (cf. *peak* vs. *peel*, in which the final phones differ in place of articulation, manner of articulation and voicing). However, further analyses showed that the effect of PND on picture naming reaction times was not significant on either words that have a close initial voiced minimal pair or words that do not have a close minimal pair, suggesting that there was no interaction effect between PND and MP condition on reaction time (Figure 5). On the other hand, since all of the words in the current experiments start with a limited selection of onsets (i.e., /p, t, k/) and have similar characteristics in general, there may not have been enough variance in word reading times to reveal a significant effect of phonological neighborhood density (As seen in Figure 1b; 469.1 – 547.5 ms). Additionally,

as a special form of phonological neighbors, the effect of the minimal pair condition on word retrieval times was not significant in either picture naming or word naming, consistent with previous studies in word naming (e.g., Peramunage et al., 2011). However, when combining the two paradigms, the interaction effect between MP condition and paradigm on reaction time became nearly significant before adding PND to the model. As observed in Figure 2, the effect of MP condition on reaction time was larger in picture naming compared to word naming. This interaction effect may be influenced by the variance difference in reaction times between two paradigms. Specifically, there may have been insufficient variance in word naming reaction times (item level variance = 9081.88 ms²) but sufficient variance in picture naming reaction times (item level variance = 58119.47 ms²) to elicit minimal pair effects (Levene's Test $p < .001$).

In contrast to the variable RT results, phonological neighborhood characteristics significantly affected VOTs. Across both picture naming and word naming, higher phonological neighborhood density was associated with longer VOTs. Several previous studies have reported significant effects of phonological neighborhood density on VOTs in word naming, supporting interactive processes in word production (Fox et al., 2015; Fricke et al., 2016). Specifically, these results suggest that as phonological overlap between the target and its neighbors increases, this leads to longer VOTs. The current study reported the same pattern of results in picture naming, suggesting that this effect holds in a more semantically driven task. While the effect of PND on VOT is consistent with the interactive nature of word production (i.e., the feed-back activation of phonological neighbors affected phonetic realization), our results could also potentially be explained by other accounts such as the speaker's monitoring process or the exemplar account (Pierrehumbert, 2002). For instance, according to a monitoring account, the perceptual similarity of words in dense neighborhoods motivates lexically conditioned phonetic variation (Luce & Pisoni, 1998; McMurray, Tanenhaus, & Aslin, 2002). According to the exemplar account, words in dense neighborhoods would be stored and produced with a more extreme articulation, maximally separating this word's phonetic distribution from that of neighboring words. However, previous studies have limited the possibility of these alternative accounts by testing word naming in different contexts and showing that words with minimal pairs were always produced with longer VOTs and presenting the words with their neighbors increased VOTs (Baese-Berk & Goldrick, 2009). Although the current study offers evidence in support of interactive models of language production, the precise nature of the mechanism behind it is beyond the scope of the current paper.

Critically, as a special form of phonological neighbors, the effect of having a minimal pair neighbor on VOTs was only significant in word naming – words that had an initial voiced contrasting minimal pair neighbor elicited significantly longer VOTs than words that did not have such neighbors. It is noteworthy that the effects of MP on VOTs became no longer significant after adding PND in the model, suggesting that the VOT effects that we observed from MP status potentially shared variance with PND. However, the interaction between MP status and paradigm on VOTs was nearly significant ($p = .051$), with larger effects of close phonological neighbors in word naming. Although this relationship became less but still marginally significant after excluding RT from VOT models ($p = .09$, See Supplemental Materials for details), we believe VOTs models including RT as a covariate were more

suitable given the different relationships between VOT and RT in the two paradigms (See Footnote 3 for details) and the potential carry-over effect of RT on phonetic realization (Buz & Jaeger, 2016; Fink, Oppenheim, & Goldrick, 2018). Moreover, the interaction between PND and paradigm was marginally significant ($p = .07$) such that the effect of phonological neighborhood density on VOT was slightly stronger in word naming than picture naming (Figure 3). Therefore, although the MP and PND effects may share variance, collectively, these results suggest that the effect of phonological neighbors on VOT was stronger in word naming compared to picture naming.

These results are consistent with previous studies focusing on word naming, supporting interactive models of word production (Baese-Berk & Goldrick, 2009; Peramunage et al., 2011). When comparing paradigms, we expected weaker effects of PND and MP condition in picture naming compared to word naming. This is because, in picture naming, a more semantically-mediated task, the top-down activation from semantic information to lexical retrieval may contribute most significantly to naming latency and phonetic realization. In other words, although phonological neighbors of the target word could also be activated, the activation of the target word itself should be most salient, driven by the visually-available semantic information. On the other hand, in word naming, although both feed-forward and feed-back activation are present, the word form was directly displayed to participants, which may emphasize form-based aspects of the target word and its phonological neighbors who have a very similar lexical form, in which case more subtle phonological and lexical effects may emerge.

Finally, our results speak to the effect of communication on phonetic realization (e.g., longer VOTs in voiceless stops with a voiced contrasting neighbor). If these changes in speech were merely a mechanism to maintain maximum speech clarity or a by-product of speech production habits, then robust and comparable VOT effects should have been found in both paradigms. The marginal differences in VOT effects between the two paradigms offer some evidence that the hyper-articulation effect in speech does not solely depend on speech context, which is consistent with previous studies (e.g., Baese-Berk & Goldrick, 2009).

Conclusion

Taken together, we extend previous studies demonstrating the effects of phonological neighborhood density and minimal pair condition on word retrieval times and phonetic variation. Critically, higher phonological neighborhood density was associated with longer voice onset times across both word naming and picture naming. Additionally, the existence of a contrasting initial voiced stop neighbor only significantly affected voice onset time in word production during word naming, but not in picture naming. In general, these results provide evidence in support of interactive models of word production and suggest that the speech production system dynamically adapts to the semantic, lexical, and phonological demands of a particular situation.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgment

This project was funded by R01 AG034138 from the National Institute on Aging (MTD), the Social Sciences Research Institute, and the Department of Psychology at the Pennsylvania State University. We thank Katherine Muschler, Maggie Treacy, Amanda Eads, and Maria Badanova for assistance with data collection and analyses. We also thank the staff and scientists at the Center for Language Science (CLS) for their support. The authors declare no conflicts of interest, financial or otherwise, that would preclude a fair review or publication of this manuscript.

References

- Adelman JS, & Brown GD (2007). Phonographic neighbors, not orthographic neighbors, determine word naming latencies. *Psychonomic Bulletin & Review*, 14(3), 455–459. [PubMed: 17874587]
- Andrews S (1989). Frequency and neighborhood effects on lexical access: Activation or search? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15(5), 802–814.
- Andrews S (1992). Frequency and neighborhood effects on lexical access: Lexical similarity or orthographic redundancy? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18(2), 234.
- Baese-Berk M, & Goldrick M (2009). Mechanisms of interaction in speech production. *Language and Cognitive Processes*, 24(4), 527–554. [PubMed: 19946622]
- Balota DA, Yap MJ, Hutchison KA, Cortese MJ, Kessler B, Loftis B, ... Treiman R (2007). The English lexicon project. *Behavior Research Methods*, 39(3), 445–459. [PubMed: 17958156]
- Barr DJ, Levy R, Scheepers C, & Tily HJ (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278.
- Barry C, Morrison CM, & Ellis AW (1997). Naming the Snodgrass and Vanderwart pictures: Effects of age of acquisition, frequency, and name agreement. *The Quarterly Journal of Experimental Psychology*, 50(3), 560–585.
- Bates D, Mächler M, Bolker B, & Walker S (2014). Fitting linear mixed-effects models using lme4. *Arxiv Preprint Arxiv:1406.5823*.
- Baus C, Costa A, & Carreiras M (2008). Neighbourhood density and frequency effects in speech production: A case for interactivity. *Language and Cognitive Processes*, 23(6), 866–888.
- Boersma P, & Weenink D (2002). Praat 4.0: a system for doing phonetics with the computer [Computer software]. Amsterdam: Universiteit Van Amsterdam.
- Brodeur MB, Guérard K, & Bouras M (2014). Bank of Standardized Stimuli (BOSS) phase II: 930 new normative photos. *Plos One*, 9(9), e106953. [PubMed: 25211489]
- Burke DM, MacKay DG, Worthley JS, & Wade E (1991). On the tip of the tongue: What causes word finding failures in young and older adults? *Journal of Memory and Language*, 30(5), 542–579.
- Burke DM, & Shafto MA (2008). *Language and aging* (Craig F & Salthouse T Eds.). New York: Psychology Press.
- Buz E, & Jaeger TF (2016). The (in) dependence of articulation and lexical planning during isolated word production. *Language, Cognition and Neuroscience*, 31(3), 404–424.
- Buz E, Tanenhaus MK, & Jaeger TF (2016). Dynamically adapted context-specific hyper-articulation: Feedback from interlocutors affects speakers' subsequent pronunciations. *Journal of Memory and Language*, 89, 68–86. [PubMed: 27375344]
- Carroll JB, & White MN (1973). Word frequency and age of acquisition as determiners of picture-naming latency. *The Quarterly Journal of Experimental Psychology*, 25(1), 85–95.
- Dell GS (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, 93(3), 283–321. [PubMed: 3749399]
- Dell GS, & O'Seaghdha PG (1992). Stages of lexical access in language production. *Cognition*, 42(1–3), 287–314. [PubMed: 1582160]
- Dell GS, Schwartz MF, Martin N, Saffran EM, & Gagnon DA (1997). Lexical access in aphasic and nonaphasic speakers. *Psychological Review*, 104(4), 801–838. [PubMed: 9337631]
- Diaz MT, Johnson MA, Burke DM, & Madden DJ (2014). Age-related differences in the neural bases of phonological and semantic processes. *Journal of Cognitive Neuroscience*, 26(12), 2798–2811. [PubMed: 24893737]

- Fink A, Oppenheim GM, & Goldrick M (2018). Interactions between lexical access and articulation. *Language, Cognition and Neuroscience*, 33(1), 12–24.
- Fox NP, Reilly M, & Blumstein SE (2015). Phonological neighborhood competition affects spoken word production irrespective of sentential context. *Journal of Memory and Language*, 83, 97–117. [PubMed: 26124538]
- Fricke M, Baese-Berk MM, & Goldrick M (2016). Dimensions of similarity in the mental lexicon. *Language, Cognition and Neuroscience*, 31(5), 639–645.
- Goldrick M (2006). Limited interaction in speech production: Chronometric, speech error, and neuropsychological evidence. *Language and Cognitive Processes*, 21(7–8), 817–855.
- Koo TK, & Li MY (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15(2), 155–163. [PubMed: 27330520]
- Levelt WJ (1999). Models of word production. *Trends in Cognitive Sciences*, 3(6), 223–232. [PubMed: 10354575]
- Levelt WJ, Roelofs A, & Meyer AS (1999). Multiple perspectives on word production. *Behavioral and Brain Sciences*, 22(01), 61–69.
- Luce PA, & Pisoni DB (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*, 19(1), 1–36. [PubMed: 9504270]
- Martin RC (2003). Language processing: functional organization and neuroanatomical basis. *Annual Review of Psychology*, 54(1), 55–89.
- McMurray B, Tanenhaus MK, & Aslin RN (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition*, 86(2), B33–B42. [PubMed: 12435537]
- Mirman D, Kittredge AK, & Dell GS (2010). Effects of near and distant phonological neighbors on picture naming. Paper presented at the Proceedings of the Annual Meeting of the Cognitive Science Society.
- Moreno-Martínez FJ, & Montoro PR (2012). An ecological alternative to Snodgrass & Vanderwart: 360 high quality colour images with norms for seven psycholinguistic variables. *Plos One*, 7(5), e37527. [PubMed: 22662166]
- Munson B, & Solomon NP (2004). The effect of phonological neighborhood density on vowel articulation. *Journal of Speech, Language, and Hearing Research*, 47(5), 1048–1058.
- Nelson NR, & Wedel A (2017). The phonetic specificity of competition: Contrastive hyperarticulation of voice onset time in conversational English. *Journal of Phonetics*, 64, 51–70.
- Peramunage D, Blumstein SE, Myers EB, Goldrick M, & Baese-Berk M (2011). Phonological neighborhood effects in spoken word production: An fMRI study. *Journal of Cognitive Neuroscience*, 23(3), 593–603. [PubMed: 20350185]
- Pierrehumbert J (2002). Word-specific phonetics. *Laboratory Phonology*, 7, 101–139.
- R Core Team. (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria 2013: ISBN 3*900051-07-0.
- Rapp B, & Goldrick M (2000). Discreteness and interactivity in spoken word production. *Psychological Review*, 107(3), 460–499. [PubMed: 10941277]
- Rizio AA, Moyer KJ, & Diaz MT (2017). Neural evidence for phonologically based language production deficits in older adults: An fMRI investigation of age-related differences in picture-word interference. *Brain and Behavior*, 7(4), e00660. [PubMed: 28413708]
- Sadat J, Martin CD, Costa A, & Alario F-X (2014). Reconciling phonological neighborhood effects in speech production through single trial analysis. *Cognitive Psychology*, 68, 33–58. [PubMed: 24291531]
- Scarborough R (2013). Neighborhood-conditioned patterns in phonetic detail: Relating coarticulation and hyperarticulation. *Journal of Phonetics*, 41(6), 491–508.
- Scarborough R, & Zellou G (2012). Perceiving Listener-directed Speech: Effects of Authenticity and Lexical Neighborhood Density. Paper presented at the Thirteenth Annual Conference of the International Speech Communication Association.
- Scarborough R, & Zellou G (2013). Clarity in communication: “Clear” speech authenticity and lexical neighborhood density effects in speech production and perception. *Journal of The Acoustical Society of America*, 134(5), 3793–3807. [PubMed: 24180789]

- Schertz J (2013). Exaggeration of featural contrasts in clarifications of misheard speech in English. *Journal of Phonetics*, 41(3–4), 249–263.
- Schwartz MF, Dell GS, Martin N, Gahl S, & Sobel P (2006). A case-series test of the interactive two-step model of lexical access: Evidence from picture naming. *Journal of Memory and Language*, 54(2), 228–264.
- Shibahara N, Shibahara N, Zorzi M, Zorzi M, Hill MP, Wydell T, & Butterworth B (2003). Semantic effects in word naming: Evidence from English and Japanese Kanji. *The Quarterly Journal of Experimental Psychology: Section A*, 56(2), 263–286.
- Snodgrass JG, & Vanderwart M (1980). A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, 6(2), 174–215. [PubMed: 7373248]
- Strain E, Patterson K, & Seidenberg MS (1995). Semantic effects in single-word naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(5), 1140–1154.
- Vitevitch MS (2002). The influence of phonological similarity neighborhoods on speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(4), 735–747.
- Vitevitch MS, & Luce PA (2004). A web-based interface to calculate phonotactic probability for words and nonwords in English. *Behavior Research Methods, Instruments, & Computers*, 36(3), 481–487.
- Vitevitch MS, & Stamer MK (2006). The curious case of competition in Spanish speech production. *Language and Cognitive Processes*, 21(6), 760–770. [PubMed: 18802492]
- Wedel A, Nelson N, & Sharp R (2018). The phonetic specificity of contrastive hyperarticulation in natural speech. *Journal of Memory and Language*, 100, 61–88.
- Wright R (2004). Factors of lexical competition in vowel articulation. *Papers in Laboratory Phonology VI*, 75–87.

Figure 1.

a) represents the relationship between phonological neighborhood density and reaction time in Picture Naming (Experiment 1); b) represents the relationship between phonological neighborhood density and reaction time in Word Naming (Experiment 2).

Figure 2. Effects of minimal pair condition on RTs in a) Picture Naming (Experiment 1); b) Word Naming (Experiment 2). Means and error bars were calculated based on participant level data.

Figure 3.

a) represents the relationship between phonological neighborhood density and VOT in Picture Naming (Experiment 1); b) represents the relationship between phonological neighborhood density and VOT in Word Naming (Experiment 2).

Figure 4. Effects of minimal pair condition on VOTs in a) Picture Naming (Experiment 1); b) Word Naming (Experiment 2). Means and error bars were calculated based on participant level data.

Figure 5.

Effect of phonological neighborhood density on picture naming reaction times in words that have a close minimal pair and words that do not have such a pair.