NOT ALL PHONOLOGICAL NEIGHBORS AFFECT PRODUCTION EQUIVALENTLY: PREDICTIONS FROM A NEURAL DYNAMIC MODEL

Michael C. Stern & Jason A. Shaw

Yale University
michael.stern@yale.edu, jason.shaw@yale.edu

ABSTRACT

Phonological neighborhood density (PND) has been shown to have apparently inconsistent effects on speech production: facilitation, inhibition, hypoarticulation, hyperarticulation, or no effect. We propose that this variety of results is due to the coarseness of the PND measure. We present a neural dynamic model of word pronunciation planning in which phonological neighbors influence phonetic target planning via a combination of excitatory and inhibitory interactions. We demonstrate through simulations how the model predicts different effects of PND depending on the identities of the phonetic dimensions differentiating neighbors. We outline novel predictions of the model that can help guide future empirical work.

Keywords: phonological neighborhoods; hyper-articulation; speech planning; Dynamic Field Theory

1. INTRODUCTION

A number of studies have investigated the effect of phonological neighborhood density (PND) on speech production, finding mixed results, e.g. facilitation [1], [2], inhibition [3], [4], hypoarticulation [5], hyperarticulation [6]–[11], or no effect [12]. The PND of a word is defined in these studies as the number of phonological neighbors in the lexicon, where a phonological neighbor is any word that differs by the addition, deletion, or substitution of a single phoneme [13]. Recently proposed models of phonetic planning based on Dynamic Field Theory (DFT) [14] have derived effects of lexical competitors on speech production, like minimal-pair-induced “contrastive hyperarticulation” [15] and “phonetic trace effects” in speech errors [16]. These lexical competition effects are similar to those that have been reported to be conditioned by PND. To probe PND effects, we present a new model that explicitly couples lexical planning to phonetic target planning within the framework of DFT. In doing so, we shed light on the existing diversity of results regarding PND effects in speech production, and generate new testable predictions.

2. MODEL STRUCTURE

A supplemental document with model equations and descriptions, as well as MATLAB scripts for running the simulations reported below, are available on OSF at https://osf.io/kc7bp/. In the model, neural populations governing the activation of lexical items are dynamically coupled to neural populations governing phonetic planning. Patterns of activation within each population unfold over time according to differential equations defined in DFT [14]. We focus on the planning of three phonetic dimensions relevant to the initial consonant: voice onset time (VOT; voicing), constriction location (CL; place of articulation), and constriction degree (CD; manner of articulation). Following [17], we define CL as a single dimension, distinct from the articulator(s) used to make the constriction. The dynamics of phonetic planning—modeled with dynamic neural fields (DNFs)—are described in [15], [16]. The novel mechanisms introduced in this paper allow dynamic coupling of lexical items to phonetic DNFs. Quite straightforwardly, a target lexical item excites its corresponding phonetic DNFs. Our proposal is that neighbors inhibit just those phonetic DNFs that differentiate them from the target. We implement this proposal with a novel mechanism for lexical-phonetic interaction that dictates the polarity of competitor activation based on the current state of phonetic DNFs.

Like antecedent DFT models [15], [16], [18]–[20], the coupling between a lexical item and a phonetic DNF is defined as a *distribution*, such that an active lexical item beginning with, e.g., a voiceless consonant, projects a distribution of activation to the VOT DNF centered on higher VOT values. Each coupling is defined as a normal distribution with position p (mean) and width w (standard deviation).

Lexical items are modeled as dynamic nodes (no internal feature gradience) with linear activation dynamics (no internal interaction). The rate of change of lexical activation is negatively related to current activation, defining point attractor dynamics. The location of the point attractor is determined by input to the lexical node (e.g., from conceptual-intentional planning, or from feedback from phonetic planning, as discussed below) and noise.

Lexical items receive feedback from phonetic DNFs to which they are coupled. This feedback causes phonological neighbors to become somewhat active during planning of a target word, and also facilitates activation of the target word. The magnitude of input from each DNF neuron to a coupled lexical item is a sigmoidal function of the neuron's current activation, such that relatively inactive neurons contribute virtually no input to coupled lexical nodes, while neurons with activation exceeding a threshold contribute substantially. Input from phonetic DNF neurons to coupled lexical nodes is modulated by the metric distance between the neuron's position in the DNF and the center of the coupling distribution p. Active neurons with a metric position close to the center of the coupling distribution (within one w of p) contribute substantially, while neurons with a position that exceeds one w from p contribute exponentially less. Crucially, the threshold for phonetic-to-lexical input is lower than the threshold for within-DNF lateral interaction (and therefore, activation peak stabilization), allowing phonological neighbors to become active before phonetic target planning is complete.

Active lexical nodes project input to phonetic DNFs to which they are coupled. The shape of this input is defined by the coupling distribution (p and w). The magnitude of this input is a positive linear function of lexical activation, such that more active lexical items project stronger input to coupled phonetic DNFs compared to less active lexical items. The polarity of this input (excitatory or inhibitory) is modulated by δ: the difference between the current state of the DNF and the lexical item's "preferred" state of the DNF, i.e., the state that would be induced by excitatory projection from the lexical item to the DNF. Polarity is a sigmoidal function of δ. When δ is relatively small (i.e., the DNF does not differ substantially from the lexical item's "preferred" state), input is excitatory. When δ exceeds a threshold (i.e., the DNF is in a "dispreferred" state for the lexical item), input from the lexical item becomes inhibitory. Crucially, the polarity of input is independent of the magnitude of input. A very active lexical item will project strong input to coupled DNFs; whether this input is excitatory or inhibitory is determined by the state of the DNF at any given time, as indexed by δ. This innovation derives the inhibitory effect of neighbors on phonetic dimensions that differ from the target.

3. SIMULATION RESULTS

We simulate initial consonant planning for a target word beginning with /t/ in three conditions, summarized in Table 1.

	Target	**Comp. 1**	**Comp. 2**
Cond. 1 "2-voiced"	/t…/ 'teen'	/d…/ 'dean'	/b…/ 'bean'
Cond. 2 "1-voiced"	/t…/ 'ten'	/d…/ 'den'	/p…/ 'pen'
Cond. 3 "0-voiced"	/t…/ 'tin'	/p…/ 'pin'	/k…/ 'kin'

Table 1: Initial consonant of the target word and each competitor in each simulation condition. Example words are given in quotes.

In each condition, the target word has two phonological neighbors differing only in the initial consonant. Thus, PND is the same in all three conditions. However, the conditions differ with respect to the phonetic dimensions differentiating the neighbors. In the 2-voiced condition, both neighbors differ from the target in initial consonant voicing (beginning with /d/ or /b/). In the 1-voiced condition, one neighbor differs from the target in initial consonant voicing (beginning with /d/), but the other neighbor overlaps with the target on initial consonant voicing (beginning with /p/). In the 0-voiced condition, both neighbors overlap with the target on initial consonant voicing (beginning with /p/ or /k/). In all three conditions, all neighbors begin with stops, so they all overlap on the CD dimension. The CD DNF thus serves as a proxy for the dimensions that are shared between all neighbors, allowing some activation to spread to all neighbors.

In all simulations, the target lexical item receives positive input, representing an intention to say that word. No other lexical item, and no phonetic DNF, receives direct input; rather, they receive activation only via interactions within the system. Figure 1 shows the timecourse of activation of the three lexical items (top) and the VOT DNF (bottom) for a single run in each condition. In all conditions, the two phonological neighbors become active via feedback from phonetic dimensions on which they overlap with the target. In the 2-voiced condition (left), both neighbors project inhibitory input to the voiced side of the VOT DNF, because by the time the neighbors start receiving activation, the state of the VOT DNF is more consistent with a voiceless consonant than a voiced consonant, resulting in a high value of δ for the neighbors. Inhibitory input to the voiced side of the DNF causes a rightward shift in the location of the activation peak, i.e., a hyperarticulated voiceless target. In the 1-voiced condition (center), only the neighbor beginning with /d/ projects inhibitory input to the voiced side of the DNF, so the rightward peak shift is smaller (hyperarticulation is decreased). Finally, in the 0-voiced condition, there is no inhibitory input to the VOT DNF, and thus no hyperarticulation. By-condition differences in the

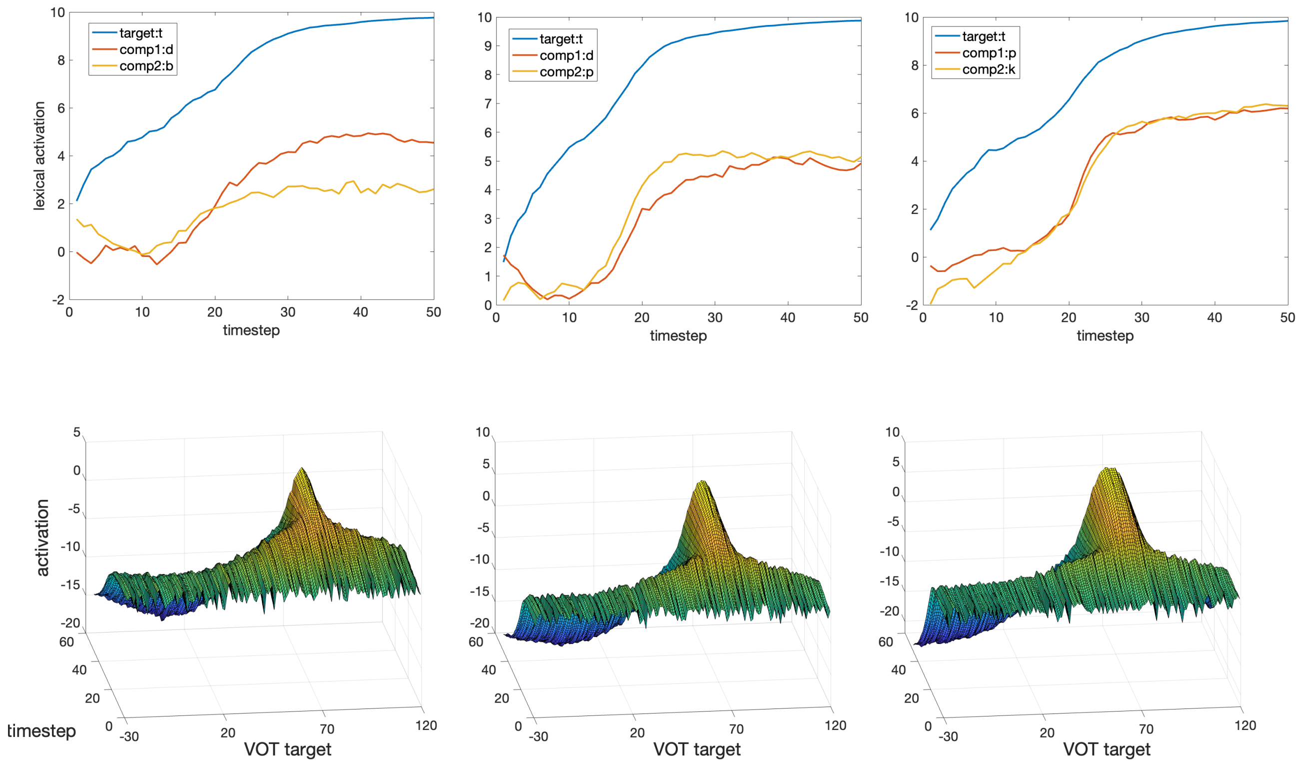

Figure 1. Lexical node (top) and VOT DNF (bottom) activation over time, single run. Left: two inhibitory competitors (2-voiced). Center: One inhibitory, one excitatory competitor (1-voiced). Right: two excitatory competitors (0-voiced).

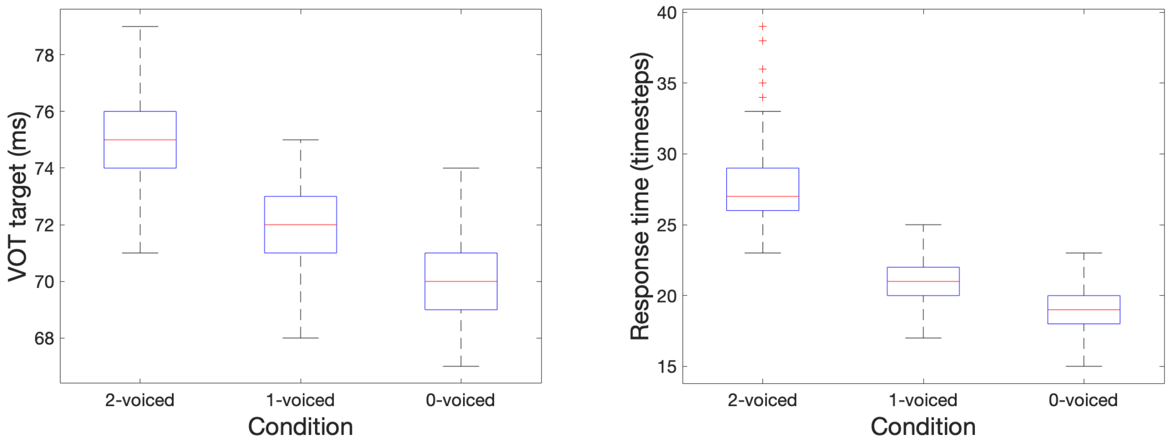

Figure 2. VOT target (left) and time to plan VOT target (right) in 500 simulated utterances in each condition.

time until an activation peak stabilizes can also be seen: slowest in the 2-voiced condition, followed by 1-voiced, followed by 0-voiced.

To investigate the systematicity of these patterns, we simulated 500 productions in each of the three conditions and recorded the VOT target (metric position of the VOT DNF neuron with maximum activation at the final timestep) and response time (peak of the first derivative of total lateral interaction) for each simulation. The target word always projected input to the VOT DNF centered at p = 70. As seen in Figure 2, a cline of VOT hyperarticulation was observed, with the most hyperarticulation in the 2-voiced condition, followed by 1-voiced, followed by 0-voiced. An analogous pattern was observed in response time.

4. DISCUSSION

Our model of the neural dynamics of lexical and phonetic planning demonstrated that different kinds of phonological neighborhoods are predicted to have different effects on articulation and planning time, even when PND is equal. The model thus offers a possible explanation underlying the existing diversity of results regarding the effects of PND on speech production, and offers testable predictions to guide future empirical work. As a first step, we expect that a better predictor than PND would be the *ratio of inhibitory to excitatory neighbors* on a given dimension. This ratio is predicted to positively correlate with the magnitude of hyperarticulation on that dimension, and the time it takes to plan articulation on that dimension. Another prediction relates to the lexical-phonetic coupling widths *w*. In the simulations, VOT was hyperarticulated due to overlap from wide inputs $w = 30$ ms. We assume that input widths transparently reflect the distribution of the phonetic cue in the listener's environment. This assumption is based on the finding that a VOT input width of 30 ms—the approximate standard deviation of VOT in American English stops [21]—derives contrastive hyperarticulation in non-errors [15] and trace effects in errors [16] of empirically observed magnitudes. We thus predict that overlap of phonetic cues in corpus distributions will correlate with the magnitude of neighborhood-induced hyperarticulation.

A third prediction relates to the feedback mechanism which activates phonological neighbors. It can be seen in Figure 1 (top, center) that the competitor beginning with /p/ received slightly more activation than the competitor beginning with /d/, even though both /d/ and /p/ differ from /t/ in only a single feature. This is because the VOT distribution (which /t/ and /p/ share) is wider ($w = 30$ ms) than the CL distribution ($w = 10$ mm), which /t/ and /d/ share. Since phonetic-to-lexical input is sensitive to *w*, the VOT DNF ends up contributing more activation to coupled lexical items than the CL DNF. This predicts that corpus variability on a phonetic dimension will correlate with the degree of excitatory lexical interaction between neighbors that overlap on that dimension, measurable via, e.g., priming or error rate. Broadly, this suggests that not only the number of overlapping phonetic dimensions, but also the way that neighbors are coupled to those dimensions, shapes the outcome of lexical interaction (cf. [22], [23]). The last prediction that we point out is related to the width of the activation peaks. It can be seen in Figure 1 (bottom) that activation peaks tend to be wider when there are more excitatory than inhibitory neighbors. The model thus predicts that words with more phonological neighbors overlapping on a particular phonetic dimension will tend to be *more variable* on that dimension.

Before concluding, we briefly discuss a few ways that this model should continue to be expanded. In our simulations, we took the time for an activation peak to stabilize in the VOT DNF as a proxy for time to speech initiation. However, all dimensions relevant to a consonant, or perhaps the whole syllable or set of upcoming syllables, likely interact in determining the time it takes to begin speaking [24]. While we expect that VOT planning time should correlate with response time, a more complete model would incorporate additional phonetic dimensions when relating model behavior to human response times. This relates to a second avenue for model development. In the current model, the neural processes governing lexical and phonetic planning unfold over time. However, the temporal coordination of phonetic dimensions is not planned. Like other models of PND in speech production (e.g., [25]), all phonetic dimensions are planned simultaneously, with no mechanism to sequence their articulation. This is likely insufficient, since the position in the word at which phonological neighbors differ has been found to influence lexical interaction [7]. The model should be elaborated to include planning of temporal coordination, while capturing these empirical effects.

5. CONCLUSION

We presented a neural dynamic model of word pronunciation planning based on Dynamic Field Theory (DFT). The model expands on previous related models by adding lexical items, and formalizing dynamic coupling mechanisms between lexical items and dimensions of phonetic planning. These coupling mechanisms account for activation of phonological neighbors during target word planning, as well as the influence (inhibitory or excitatory) of phonological neighbors on phonetic planning. We demonstrated through simulations that the ratio of phonological neighbors that differ on a phonetic dimension ("inhibitory" neighbors) to neighbors that overlap on a phonetic dimension ("excitatory" neighbors) is expected to be a better predictor of hyperarticulation on that dimension than global phonological neighborhood density (PND), offering a possible explanation of the existing diversity of results, as well as predictions to guide future empirical work.

As a final point, the flow of activation in the model (both laterally within DNFs, and between lexical items and DNFs) is highly non-linear, allowing stabilization over time on a particular word's pronunciation plan. However, the primitive features of the model are all continuous, allowing the generation of gradient shifts in pronunciation and planning time, even in the face of the categoricity generated by the non-linearities. A general strength of this kind of modeling approach is that gradience and categoricity co-exist in the same dimensions.

6. REFERENCES

[1] M. S. Vitevitch, “The Influence of Phonological Similarity Neighborhoods on Speech Production,” *J. Exp. Psychol. Learn. Mem. Cogn.*, vol. 28, no. 4, pp. 735–747, 2002, doi: 10.1037/0278-7393.28.4.735.

[2] E. Buz and T. F. Jaeger, “The (in)dependence of articulation and lexical planning during isolated word production,” *Lang. Cogn. Neurosci.*, vol. 31, no. 3, pp. 404–424, 2016, doi: 10.1080/23273798.2015.1105984.

[3] A. F. Frank, M. K. Tanenhaus, R. N. Aslin, and A. P. Salverda, “Frequency, neighborhood density, and phonological similarity effects in picture naming: An artificial lexicon study,” *Proc. 29th Annu. Cogn. Sci. Soc.*, pp. 995–1000, 2007.

[4] J. K. Gordon and J. C. Kurczek, “The aging neighborhood: phonological density in naming,” *Lang. Cogn. Neurosci.*, vol. 29, no. 3, pp. 326–344, Mar. 2014, doi: 10.1080/01690965.2013.837495.

[5] S. Gahl, Y. Yao, and K. Johnson, “Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech,” *J. Mem. Lang.*, vol. 66, no. 4, pp. 789–806, 2012, doi: 10.1016/j.jml.2011.11.006.

[6] N. P. Fox, M. Reilly, and S. E. Blumstein, “Phonological neighborhood competition affects spoken word production irrespective of sentential context,” *J. Mem. Lang.*, vol. 83, pp. 97–117, 2015, doi: 10.1016/j.jml.2015.04.002.

[7] M. Fricke, M. M. Baese-Berk, and M. Goldrick, “Dimensions of similarity in the mental lexicon,” *Lang. Cogn. Neurosci.*, vol. 31, no. 5, pp. 639–645, 2016, doi: 10.1080/23273798.2015.1130234.

[8] M. Fricke, “Phonological enconding and phonetic duration,” Doctoral dissertation, University of California, Berkeley, 2013.

[9] B. Munson and N. P. Solomon, “The effect of phonological neighborhood density on vowel articulation,” *J. Speech, Lang. Hear. Res.*, vol. 47, no. 5, pp. 1048–1058, 2004, doi: 10.1044/1092-4388(2004/078).

[10] B. Munson, “Lexical access, lexical representation, and vowel production,” *Lab. Phonol.*, vol. 9, pp. 201–228, 2007.

[11] R. Scarborough, “Neighborhood-conditioned patterns in phonetic detail: Relating coarticulation and hyperarticulation,” *J. Phon.*, vol. 41, no. 6, pp. 491–508, 2013, doi: 10.1016/j.wocn.2013.09.004.

[12] J. J. Holliday and R. Turnbull, “Effects of Phonological Neighborhood Density on Word Production in Korean,” *Proc. 18th Int. Congr. Phonetic Sci. (ICPhS 2015)*, pp. 1–5, 2015.

[13] P. A. Luce, “Neighborhoods of Words in the Mental Lexicon,” Doctoral dissertation, Indiana University, 1986.

[14] G. Schöner, J. Spencer, and D. R. Group, *Dynamic Thinking: A Primer on Dynamic Field Theory*. Oxford: Oxford University Press, 2016.

[15] M. C. Stern and J. A. Shaw, “Neural inhibition during speech planning contributes to contrastive hyperarticulation,” *arXiv Prepr. arXiv2209.12278*, 2022, doi: https://doi.org/10.48550/arXiv.2209.12278.

[16] M. C. Stern, M. Chaturvedi, and J. A. Shaw, “A dynamic neural field model of phonetic trace effects in speech errors,” *Proc. Annu. Meet. Cogn. Sci. Soc. 44*, 2022.

[17] C. P. Browman and L. Goldstein, “Articulatory gestures as phonological units,” *Phonology*, vol. 6, no. 2, pp. 201–251, 1989, doi: 10.1017/S0952675700001019.

[18] C. Kirov and A. I. Gafos, “Dynamic phonetic detail in lexical representations,” *ICPhS XVI*, pp. 637–640, 2007.

[19] A. Gafos and C. Kirov, “A dynamical model of change in phonological representations: The case of lenition,” *Approaches to Phonol. Complex.*, pp. 219–240, 2009, doi: 10.1515/9783110223958.219.

[20] K. D. Roon and A. I. Gafos, “Perceiving while producing: Modeling the dynamics of phonological planning,” *J. Mem. Lang.*, vol. 89, pp. 222–243, 2016, doi: 10.1016/j.jml.2016.01.005.

[21] E. Chodroff and C. Wilson, “Structure in talker-specific phonetic realization: Covariation of stop consonant VOT in American English,” *J. Phon.*, vol. 61, pp. 30–47, 2017, doi: 10.1016/j.wocn.2017.01.001.

[22] P. C. Gordon and D. E. Meyer, “Perceptual-motor processing of phonetic features in speech,” *J. Exp. Psychol. Hum. Percept. Perform.*, vol. 10, no. 2, pp. 153–178, 1984, doi: 10.1037/0096-1523.10.2.153.

[23] A. Ulicheva, K. D. Roon, Z. Cherkasova, and P. Mousikou, “Effects of phonological features on reading-aloud latencies: A cross-linguistic comparison,” *J. Exp. Psychol. Learn. Mem. Cogn.*, vol. 48, no. 9, pp. 1348–1362, 2021, doi: 10.1037/xlm0000893.

[24] E. Bates *et al.*, “Timed picture naming in seven languages,” *Psychon. Bull. Rev.*, vol. 10, no. 2, pp. 344–380, 2003, doi: 10.3758/BF03196494.

[25] P. A. Luce and D. B. Pisoni, “Recognizing spoken words: the neighborhood activation model.,” *Ear Hear.*, vol. 19, no. 1, pp. 1–36, Feb. 1998, doi: 10.1097/00003446-199802000-00001.