Neural inhibition during speech planning contributes to contrastive hyperarticulation

Michael C. Stern*, Jason A. Shaw

Department of Linguistics, Yale University, 370 Temple St, New Haven, CT, USA

ARTICLE INFO

Keywords:
Contrastive hyperarticulation
Inhibition
Dynamic field theory
Phonological neighborhoods
Voice onset time

ABSTRACT

Previous work has demonstrated that words are hyperarticulated on dimensions of speech that differentiate them from a minimal pair competitor. This phenomenon has been termed contrastive hyperarticulation (CH). We present a dynamic neural field (DNF) model of voice onset time (VOT) planning that derives CH from an inhibitory influence of the minimal pair competitor during planning. We test some predictions of the model with a novel experiment investigating CH of voiceless stop consonant VOT in pseudowords. The results demonstrate a CH effect in pseudowords, consistent with a basis for the effect in the real-time planning and production of speech. The scope and magnitude of CH in pseudowords was reduced compared to CH in real words, consistent with a role for interactive activation between lexical and phonological levels of planning. We discuss the potential of our model to unify an apparently disparate set of phenomena, from CH to phonological neighborhood effects to phonetic trace effects in speech errors.

Introduction

Contrastive hyperarticulation

Previous work has shown that the pronunciation of a word is gradually influenced by its relationships with other similar words in the lexicon. For example, phonological neighborhood density (Luce, 1986; Vitevitch & Luce, 2016)—defined broadly as the number of similar-sounding words in the lexicon (Vitevitch & Luce, 2016)—has been shown to affect the pronunciation of both vowels (Gahl et al., 2012; Munson, 2007; Munson & Solomon, 2004, 2016; Scarborough, 2010, 2012, 2013; Wright, 2004) and consonants (Fox et al., 2015; Fricke et al., 2016). The number of similar-sounding words in the lexicon seems to influence pronunciation. Phonological neighborhood effects on pronunciation can be characterized as global, since they are not targeted towards specific lexical competitors. However, a more local effect is caused by the presence of a minimal pair competitor. We consider a minimal pair competitor to be a word that differs from the target word in a single dimension of speech or “phonological feature”. The main phonological feature of interest in this paper is “voicing”, which differentiates, e.g., the initial consonant of PET from that of minimal competitor BET. Consonant voicing corresponds articulatorily to the temporal coordination between oral and laryngeal gestures; voicing can be measured acoustically as the duration of the interval between the “burst” caused by the sudden release of an oral occlusion (e.g., of the lips for /p/) and the onset of periodic vocal fold vibration associated with the following vowel. This duration is termed “voice onset time” (VOT: Abramson & Whalen, 2017; Lisker & Abramson, 1964). In conversational speech, “voiceless” consonants like /p, t, k/ are defined by a long VOT (about 40–80 ms), while “voiced” consonants like /b, d, g/ are defined by a short VOT (about 10–30 ms) (e.g., Chodoroff & Wilson, 2017).

Change in the pronunciation of a word that specifically differentiates it from a minimal pair competitor—termed “contrastive hyperarticulation” (CH: Wedel et al., 2018)—has been observed in several specific dimensions of speech, including vowel formants (Clopper & Tamarit, 2014; Wedel et al., 2018), vowel duration (Goldrick et al., 2013; Schertz, 2013; Seyfarth et al., 2016), and voicing of voiceless consonants (Baese-Berk & Goldrick, 2009; Buz et al., 2016; Goldrick et al., 2013; Nelson & Wedel, 2017; Schertz, 2013; Wedel et al., 2018) and voiced consonants (Nelson & Wedel, 2017; Schertz, 2013; Seyfarth et al., 2016; Wedel et al., 2018). CH is magnified when the minimal pair is

* Corresponding author.
E-mail address: michael.stern@yale.edu (M.C. Stern).

1 All caps indicates a lexical representation.
2 slashes indicate a phonological representation.

https://doi.org/10.1016/j.jml.2023.104443
Received 24 September 2022; Received in revised form 7 June 2023; Accepted 8 June 2023
Available online 17 June 2023
0749-596X/© 2023 Elsevier Inc. All rights reserved.
contextually salient (Baese-Berk & Goldrick, 2009; Kirov & Wilson, 2012; Seyfarth et al., 2016) and after the listener indicates that they heard the minimal pair competitor rather than the target word (Buz et al., 2016; Schertz, 2013). CH may underlie the influence of functional load—operationalized as the number of lexical items distinguished by a phonological feature (Hockett, 1967)—on articulation (Hall et al., 2017). This influence has been argued to maintain the distinctiveness of words over time (Wedel, 2012; Wedel et al., 2013; Winter & Wedel, 2016). While some have argued that global neighborhood effects are in fact reducible to local CH effects (Buz & Jaeger, 2016; Nelson & Wedel, 2017; Wedel et al., 2018) or vice versa (Fox et al., 2015; Frick et al., 2016), others have argued that the two types of effects arise from (at least partially) independent sources (Clopper & Tamati, 2014). The focus of the present study is CH, rather than general neighborhood effects; however, the results have implications for understanding the relationship between these types of effects, a point we return to in the discussion section.

Baese-Berk & Goldrick (2009: Experiment 2) offers a representative example of an experiment demonstrating CH. During each trial in this experiment, participants were shown a screen with three words, one of which (the target) was highlighted. The participant was instructed to read the target word to a “listener” (a lab confederate) who was viewing a different screen, identical to the participant’s but without the target word highlighted. The participant’s goal was to get the listener to select the correct target word. All target words were real words beginning with voiceless stop consonants (e.g., /p, t, k/), but they were divided into three conditions. Words in the “no competitor” condition had no minimal pair competitor differing in initial consonant voicing, e.g. PIPE (*BPIPE), and the two distractor words were unrelated to the target. Words in the “no context” condition had a minimal pair competitor, e.g. PAD (BAD), but this minimal pair was not presented as a distractor; again, the two distractors were unrelated to the target. In the “context” condition, the target’s minimal pair competitor was presented as an on-screen distractor. The VOT of words in the “no context” condition was hyperarticulated compared to those in the “no competitor” condition by an average of about 5 ms, phonetically differentiating the voiceless target from the voiced minimal pair competitor. Moreover, words in the “context” condition were hyperarticulated compared to those in the “no competitor” condition by about 10 ms, demonstrating an additional effect of the contextual salience of the minimal pair competitor.

Possible sources of contrastive hyperarticulation

A number of hypotheses have been proposed to explain CH. On one hand, CH may arise from relatively slow processes occurring over the lifetimes of language users. In episodic memory models of linguistic knowledge (Goldinger, 1998; Pierrrehumbert, 2001, 2002), lexical representations are phonetically detailed and continually updated after each production or perception of the lexical item. Crucially, in many implementations of such models, only successfully categorized exemplars influence the overall lexical representation (Hay et al., 2015; Wedel, 2006). Hyperarticulated productions of words with minimal pair competitors are more likely to be misclassified than those of words without minimal pair competitors. Thus, over time, phonetic representations of words with minimal pair competitors will tend to become more peripheral, i.e., more differentiated from their minimal pair competitor. Episodic memory accounts of CH thus provide a possible mechanism underlying the observation that phonological category oppositions which carry a greater functional load are less likely to merge over time (Wedel et al., 2013). However, this mechanism alone cannot account for the effect of the contextual salience of the minimal pair competitor, i.e. the difference between the “context” and the “no context” conditions.

Thus, there is likely a role for relatively fast processes unfolding on millisecond timescales in the planning and production of speech. For instance, CH has been attributed to real-time listener accommodation on the part of the speaker (Munson & Solomon, 2004; Wright, 2004). According to this account, speakers are sensitive to the perceptual needs of the listener, and actively adjust their pronunciation in an attempt to maximize the likelihood that they will be accurately perceived by the listener (Lindblom, 1990). More broadly, this would constitute a specific case of audience design, whereby speakers take into account the likely perceptual experience of the listener (Arnold et al., 2012; Bell, 1984; Galati & Brennan, 2010). This theoretical position has been deployed to explain pronunciation variation, as in CH, as well as other language behaviors, including word choice, lexico-syntactic patterns, and semantic-pragmatic reasoning (for a review see Jaeger & Buz, 2017).

With regard to CH, words with minimal pair competitors are more likely to be misperceived than words without minimal pair competitors (Vitevitch & Luce, 1998), so speakers might hyperarticulate dimensions of pronunciation that differentiate words from their minimal pair competitors, to accommodate the listener. Hyperarticulation would be further magnified when the minimal pair competitor is more salient in the context and therefore more confusable with the target. This account is supported by the finding that speakers actively adjust their pronunciations in response to an explicit misunderstanding on the part of the listener (Buz et al., 2016; Schertz, 2013).

CH has also been attributed to real-time speaker-internal processes, independent of the listener. This kind of hypothesis is based on “interactive activation” models of speech planning in which activation of one representational unit (e.g., lexical item or phoneme) causes partial activation of related representational units (Dell, 1986; Dell et al., 1999, 2021). For instance, consider the planning of the word PUN, schematized in an interactive activation framework in Fig. 1. First, intention to produce PUN increases activation of its lexical representation. Via connections between lexical and phonological levels of planning, activation of the lexical representation PUN increases activation of its constituent phonological representations /p/, /n/, and /n/. These lexical-phonological connections are bidirectional, so the active phonological representations send activation back to the lexical level, forming an excitatory feedback loop. Importantly, at this stage, lexical representations which overlap in their phonological representation with the target word also receive some activation via phonological-lexical feedback. The minimal pair Bun, differing from the target PUN in only one phonological feature (voicing of the initial consonant), receives a non-negligible amount of activation at this stage. Because of the bidirectional nature of interactive activation, the competitor lexical representation BUN subsequently sends some activation to the competitor phoneme /b/, which is not part of the phonological representation of the target word PUN. It has been proposed that partial activation of the competitor phoneme (in this case, /b/) drives

Fig. 1. During planning of PUN, interactive activation (double arrows) causes partial activation of the competitor phoneme /b/ (red arrow). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
from the visual context, in addition to the activation it receives from non-errorful production of BUN. Applied to CH, this mechanism appears actual mechanism by which competitor activation in planning leads to phoneme. In the
have a phoneme (Goldrick et al., 2016). For instance, if BUN were produced derived from different magnitudes of activation of the competitor CH could derive from partial activation of competitors, the difference interactions, and noise. An important characteristic of DNFs is their capacity to model and experimental results and suggesting directions for future work. Section ‘Conclusion’ concludes.

Dynamic neural field model of VOT planning

In DFT, features relevant to perception, behavior, and cognition are modeled as continuous parameters represented by populations of neurons. The distribution of activation (spike rate) across the neurons in a population is modeled as a dynamic neural field (DNF). DNFs evolve over time under the influence of input (e.g., from sensory surfaces), lateral interactions, and noise. An important characteristic of DNFs is their ability to resolve competition. A single DNF can be simultaneously influenced by multiple inputs, and the dynamics of the field specify how, under the right conditions, these inputs will resolve to a single output, sometimes reflecting characteristics of both inputs. This mechanism has been shown to derive the influence of perceptual input on the initiation timing of speech movements (Room & Gafos, 2016), the influence of distractors on the location of reaching movements (Erhagen & Schöner, 2002), and the influence of phonological competitors on speech targets in errors (Stern et al., 2022).

In many DNF implementations, inputs to the field are strictly excitatory, increasing activation in a region of the field until a stable peak of activation forms in that region corresponding to, e.g., a particular percept or movement plan. However, it has been proposed that regions of DNFs can also be selectively inhibited in order to prevent formation of a stable peak in that region. This mechanism has been hypothesized to regulate attention in the presence of multiple salient percepts (Houghton & Tipper, 1994). Models of selective inhibition have been shown to derive the “negative priming” (Tipper, 1985) and “inhibition of return” (Posner & Cohen, 1984) effects observed in response-distractor tasks. Of particular relevance for CH, these models also exhibit dissimilation of movement targets from distractors in manual reaching and eye saccades (Tipper et al., 2000), as well as English vowel (Tilen, 2007, 2009) and Mandarin tone production (Tilen, 2013). In order to derive dissimilation from selective inhibition, it is assumed that different feature representations (modeled as DNF input distributions) overlap to some degree in feature space. Thus, selective inhibition of, e.g., a voiced consonant category (distributed over the lower values in a VOT planning field) causes partial inhibition of the voiceless consonant category. In particular, those neurons which are sensitive to the lowest VOT values in the voiceless distribution will be inhibited. In this way, the VOT target of a voiceless production planned during selective inhibition of the voiceless category will tend to be higher, i.e. more hyperarticulated, compared to a production planned during no selective inhibition. Selective inhibition thus offers a mechanism by which influence of a minimal pair competitor on planning can lead to CH. In particular, inhibitory projection from a voiced competitor to a VOT planning field during production of a voiceless target is expected to cause increased VOT, i.e. CH. We demonstrate this in the following subsections.

Data availability

The modeling reported below was done using the MATLAB-based software COSIVINA (Schneegans, 2021). Scripts for simulating the results below are available on OSF at https://osf.io/hz8fp/.

Model structure

The structure of the DNF model of VOT planning presented here is mostly identical to that described in (Stern et al., 2022). The model is summarized in Eq. (1):

\[
\begin{align*}
\tau \ddot{u}(x,t) &= -u(x,t) + h + s_{\text{target}}(x,t) + s_{\text{inp}}(x,t) \\
&+ \int k(x-x')g(u(x',t))dx' + q_0(x,t)
\end{align*}
\]

(1)

The key component of the model is the activation field \( u(x,t) \) defined over the VOT dimension \( x \) at each moment in time \( t \). We set the range of \( x \) to 200, representing a range of possible VOT targets (in ms). The rate of change of activation \( u(x,t) \) is inversely related to current activation \( u(x,t) \), so Eq. (1) represents a dynamical system with a point attractor at \( h + s_{\text{target}}(x,t) + s_{\text{inp}}(x,t) = \int k(x-x')g(u(x',t))dx' + q_0(x,t) \). \( \tau \) is a time constant, with higher values corresponding to slower rates of field evolution. The resting level \( h \) is assumed to be below zero for all field locations (neurons), by convention at \( -5 \). Each field input \( s_{\text{target}}(x,t) \) and \( s_{\text{inp}}(x,t) \) is represented as a separate Gaussian distribution of the form...
threshold neuron. In our model, contributes additional across-the-board inhibition from each above-

\( \sigma \) (Input parameter values).

M.C. Stern and J.A. Shaw

"regions to previous conceptualizations of speech production goals as distribution (see Fig. 4)."

where \( \tau \), \( h \), \( \beta \), \( \epsilon_{\text{exc}} \), \( \epsilon_{\text{inh}} \), \( \epsilon_{\text{glob}} \), \( \epsilon_{\text{exc}} \), \( \epsilon_{\text{inh}} \), and \( \theta \) as well as the parameters of the interaction kernel (see Fig. 2). The global inhibition parameter \( \epsilon_{\text{glob}} \) is large enough relative to the range of input amplitudes that we consider to ensure selection dynamics. That is, under these conditions, only a single peak will form regardless of the number of inputs to the field. This is crucial because human speech only allows the production of one VOT value per segment. The values of the input parameters (Table 2) were determined as follows. \( s_{\text{target}} \) and \( s_{\text{mp}} \) (the centers of the input distributions) were set to 70 ms and 20 ms, respectively, and both \( w_{\text{target}} \) and \( w_{\text{mp}} \) were set to 30, broadly consistent with recent reports of means and standard deviations measured for American English stop consonants (Chodroff & Wilson, 2017). \( a_{\text{target}} \) was set to 6 in all simulations in order to ensure the formation of a stable activation peak in the context of a resting activation level \( h = -5 \). In order to investigate the effects of influence from a voiced minimal pair on VOT planning for a voiceless target, we varied \( a_{\text{mp}} \) from -6 to 4 in steps of 0.5. At each value of \( a_{\text{mp}} \), we simulated 500 instances of field evolution with 120 time steps each. The VOT target for each simulation was calculated as the point of maximum activation in the field at the final time step. Fig. 5 displays the results.

Fig. 5 demonstrates a clear trend whereby decreasing voiced input amplitude, \( a_{\text{mp}} \) (increasing the magnitude of inhibitory input), corresponds to a larger (more hyperarticulated) voiceless VOT target. This is because \( s_{\text{target}} \) and \( s_{\text{mp}} \) partially overlap, so inhibitory input from \( s_{\text{mp}} \) depresses activation in those regions of the field that correspond to a more hyperarticulated voiceless VOT production. This is consistent with the empirical observation that CH of voiceless consonants is driven by a decrease in the likelihood of hypoarticulated productions, rather than an overall shift in the VOT distribution; i.e., CH is the result of a change in the skewness of the VOT distribution, rather than the mode (Buz et al., 2016). Increasing the amplitude of inhibitory input increases the magnitude of hyperarticulation: there is an approximately linear

### Table 1

| Parameter Value |
|-----------------|
| Value |
| \( \tau \) | 20 |
| \( h \) | -5 |
| \( \beta \) | 4 |
| \( \epsilon_{\text{exc}} \) | 15 |
| \( \epsilon_{\text{inh}} \) | 5 |
| \( \epsilon_{\text{glob}} \) | 0.9 |
| \( \epsilon_{\text{exc}} \) | 5 |
| \( \epsilon_{\text{inh}} \) | 12.5 |
| \( q \) | 1 |

### Table 2

| Input parameter values. |
|-------------------------|
| Parameter |
| Value |
| \( s_{\text{target}} \) (voiceless target) |
| \( p \) | 70 |
| \( w \) | 30 |
| \( a \) | 6 |
| \( s_{\text{mp}} \) (voiced minimal pair) |
| \( p \) | 20 |
| \( w \) | 30 |
| \( a \) | varies |

where \( a \) controls the amplitude or strength of the input, \( p \) controls the position of the input in the field, and \( w \) controls the width of the input distribution (see Fig. 4). \( s_{\text{target}} \) represents the target voiceless distribution, and \( s_{\text{mp}} \) represents the voiced minimal pair distribution. This treatment of speech intentions as distributions in feature space is similar to previous conceptualizations of speech production goals as “ranges” (Byrd & Saltzman, 2003), “windows” (Keating, 1990) or “convex regions” (Guenther, 1995). Each neuron \( x \) which exceeds an activation threshold contributes activation to other neurons \( x \) via an interaction kernel \( k(x - x^{'}) \) given by

\[
k(x - x^{'}) = \frac{c_{\text{exc}}}{\sqrt{2\pi\sigma_{\text{exc}}}} \exp \left[ -\frac{(x - p)^2}{2\sigma_{\text{exc}}^2} \right] - \frac{c_{\text{inh}}}{\sqrt{2\pi\sigma_{\text{inh}}}} \exp \left[ -\frac{(x - x^{'})^2}{2\sigma_{\text{inh}}^2} \right] - \frac{c_{\text{glob}}}{\sqrt{2\pi\sigma_{\text{glob}}}} \exp \left[ -\frac{(x - x^{'})^2}{2\sigma_{\text{glob}}^2} \right] - c_{\text{glob}}
\]

The effects of both excitatory and inhibitory interaction are modeled as Gaussian distributions centered on each neuron \( x \). \( c_{\text{exc}} \) and \( c_{\text{inh}} \) control the magnitude of excitatory and inhibitory interaction, respectively, and \( \sigma_{\text{exc}} \) and \( \sigma_{\text{inh}} \) control the width of each interaction distribution. \( c_{\text{glob}} \) contributes additional across-the-board inhibition from each above-threshold neuron. In our model, \( c_{\text{exc}} > c_{\text{inh}} > c_{\text{glob}} \) and \( \sigma_{\text{exc}} < \sigma_{\text{inh}} \). So interaction is excitatory (positive effect on activation) for nearby neurons and inhibitory (negative effect on activation) for more distant neurons. Lateral excitation contributes to the stabilization of activation peaks which drive articulation, while lateral inhibition prevents runaway expansion of activation peaks. The activation threshold for interaction is given by a sigmoidal function \( g(u) \), where \( \beta \) controls the steepness of the threshold, as seen in Eq. (4) and Fig. 3.

\[
g(u) = \frac{1}{1 + \exp(-\beta u)}
\]

By convention, the threshold is set to \( u = 0 \) so that lateral interaction kicks in only when activation is positive. Finally, noise is simulated by adding normally distributed random values \( \xi(x, t) \) weighted by a parameter \( q \).

### Simulation results

In this subsection, we use the model described above to simulate VOT planning for voiceless stop consonants in a number of conditions. The values of the field parameters used in all simulations are listed in Table 1, and the values of the input parameters are listed in Table 2. The input distributions are plotted in Fig. 4.

Since part of our aim is to unify CH with trace effects in speech errors, we set the parameters of the DNF to be identical to those in Stern et al. (2022). This includes \( \tau \), \( h \), \( \beta \), and \( q \) as well as the parameters of the interaction kernel (see Fig. 2). The global inhibition parameter \( c_{\text{glob}} \) is large enough relative to the range of input amplitudes that we consider to ensure selection dynamics. That is, under these conditions, only a single peak will form regardless of the number of inputs to the field. This is crucial because human speech only allows the production of one VOT value per segment. The values of the input parameters (Table 2) were determined as follows. \( s_{\text{target}} \) and \( s_{\text{mp}} \) (the centers of the input distributions) were set to 70 ms and 20 ms, respectively, and both \( w_{\text{target}} \) and \( w_{\text{mp}} \) were set to 30, broadly consistent with recent reports of means and standard deviations measured for American English stop consonants (Chodroff & Wilson, 2017). \( a_{\text{target}} \) was set to 6 in all simulations in order to ensure the formation of a stable activation peak in the context of a resting activation level \( h = -5 \). In order to investigate the effects of influence from a voiced minimal pair on VOT planning for a voiceless target, we varied \( a_{\text{mp}} \) from -6 to 4 in steps of 0.5. At each value of \( a_{\text{mp}} \), we simulated 500 instances of field evolution with 120 time steps each. The VOT target for each simulation was calculated as the point of maximum activation in the field at the final time step. Fig. 5 displays the results.

Fig. 5 demonstrates a clear trend whereby decreasing voiced input amplitude, \( a_{\text{mp}} \) (increasing the magnitude of inhibitory input), corresponds to a larger (more hyperarticulated) voiceless VOT target. This is because \( s_{\text{target}} \) and \( s_{\text{mp}} \) partially overlap, so inhibitory input from \( s_{\text{mp}} \) depresses activation in those regions of the field that correspond to a more hyperarticulated voiceless VOT production. This is consistent with the empirical observation that CH of voiceless consonants is driven by a decrease in the likelihood of hypoarticulated productions, rather than an overall shift in the VOT distribution; i.e., CH is the result of a change in the skewness of the VOT distribution, rather than the mode (Buz et al., 2016). Increasing the amplitude of inhibitory input increases the magnitude of hyperarticulation: there is an approximately linear
Fig. 2. Lateral interaction kernel $k(x-x')$.

Fig. 4. Right: voiceless target input distribution with constant positive amplitude $a$. Left: voiced minimal pair input distribution with varying $a$. Blue lines indicate positive values of $a$; red lines indicate negative values of $a$. The three bold lines correspond to the highlighted results in Fig. 5. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Fig. 5. VOT targets generated by the DNF under the influence of two inputs: the voiceless target $s_{\text{target}}$ and the voiced minimal pair competitor $s_{\text{mp}}$ of varying amplitude, $a_{\text{mp}}$. Dotted black line: $a_{\text{mp}} = 0$, CH = 0 (“no competitor” condition from Baese-Berk & Goldrick, 2009). Dashed red line: $a_{\text{mp}} = -3$, CH ≈ 5 (“no context” condition). Solid red line: $a_{\text{mp}} = -6$, CH ≈ 10 (“context” condition). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
correlation between $a_{\text{mp}}$ and VOT target. The range of CH magnitudes observed in Fig. 5 covers that observed in the three conditions from Baese-Berk & Goldrick (2009:Experiment 2), and allows us to understand the difference between these conditions as a difference in strength of inhibitory input from the minimal pair competitor. An example of field evolution in each of these conditions is displayed in Fig. 6.

The field location of the activation peak shifts slightly depending on the magnitude of inhibitory voiced input: 70 ms (left), 75 ms (middle), and 80 ms (right). We can also observe that inhibitory voiced input slows down the formation of an activation peak. With no inhibitory input, the first neurons surpass the threshold for interaction ($u = 0$) at about the 40th time step, and the subsequent combination of local lateral excitation and global lateral inhibition causes an activation peak to form quite quickly. With intermediate inhibition, the first neurons do not surpass the threshold until about the 60th time step, and with strong inhibition, no neurons surpass the threshold until the very end of the 120-step simulation. In fact, in some simulations, 120 time steps was not enough time for any activation peak to form when $a_{\text{mp}} = -6$. Thus, this model predicts a positive relationship between planning time and the magnitude of CH, a prediction which could be tested with an experiment designed to measure both response time and VOT.

Another observation from Fig. 5 is that, when $a_{\text{mp}}$ is positive, we observe a decrease in VOT, i.e., a hyperarticulation effect, consistent with the trace effect in speech errors (Stern et al., 2022). This framework thus allows us to understand both trace effects in speech errors and CH effects in non-errors as arising from the influence of lexical competitors on speech planning. The difference is that in the former case, this influence is excitatory, while in the latter case, it is inhibitory.

By understanding the range of previously observed CH effect magnitudes as arising from a cline of amplitudes of competitor inputs to a DNF governing pronunciation planning, we predict that intermediate input amplitudes should lead to intermediate CH effect magnitudes. For instance, VOT planning for a voiceless target under the influence of a voiced minimal pair competitor with input amplitude between 0 and –3 should lead to a small CH effect, i.e., smaller than 5 ms, the CH effect observed in the “no context” condition. One way to reduce the input amplitude of the competitor is to examine pseudoword production. Since pseudowords have no lexical representation, the network of interaction that activates the competitor has one less node, thereby decreasing the activation of the whole network, including the competitor (see Fig. 1).

In order to further vary the amplitude of competitor input, we can additionally introduce visual input to the competitor, similar to the “context” condition. The model predicts that CH should be observable for pseudowords, but it should be reduced in magnitude compared to CH in real words. We tested these predictions with an experiment, described in the following section.

### Speech production experiment design

Our experimental design is largely a replication of Baese-Berk & Goldrick (2009:Experiment 2) with two main differences: target items were pseudowords, rather than real words, and the experiment was conducted over Zoom, rather than in the lab.

#### Data availability

All figures and analyses reported below were done in R version 4.1.2 (R Core Team, 2021). The anonymized data and R code are available on OSF at https://osf.io/zz8fp/.

#### Participants

24 adults participated in the experiment (ages 18–36, $M = 24.88$, $SD = 4.97$; 17 women, 7 men). This number of participants is twice that included in Baese-Berk & Goldrick (2009:Experiment 2), and was chosen in order to maximize the likelihood of detecting an effect (minimize the likelihood of a Type II error), given an expected increase in noise in online compared to in-person data collection. All participants self-reported that they were native speakers of American English, and that they had no history of speech, language, or hearing impairment. All participants provided informed consent under Yale University IRB #2000030436.

#### Materials

The experimental stimuli consisted of mono- and di-syllabic pseudowords beginning with voiceless stop consonants (see Appendix A for the complete list of stimuli). Two independent variables were manipulated: (1) whether or not the pseudoword had a minimal pair competitor in word-initial stop consonant voicing (e.g., TIVE /tɪːv/, which has a minimal pair competitor DIVE /dɪːv/ in the lexicon, vs. TIBE /tɪːb/, which lacks a minimal pair competitor DIBE /dɪːb/ in the lexicon) and (2) whether or not the minimal pair was salient in the context, i.e., presented as a competitor on the screen or not. Crossing these factors yields four experimental conditions. There were 12 target items per condition (four items for each major place of articulation: labial (/p/), alveolar (/t/), and velar (/k/)), for a total of 48 experimental target items. For the same reasons described above for participant selection, this number of experimental target items was chosen to be a substantial
pseudowords, and, like the experimental items, varied according to unrelated to the experimental items. Filler items included both real and speaker’s independent variables. Measures are from (Vitevitch minimal pair on screen).”

Results of Welch’s t-tests comparing control variables between levels of the two independent variables. Measures are from (Vitevitch & Luce, 2004).

|                          | Minimal pair in lexicon vs. no minimal pair in lexicon | Minimal pair on screen vs. no minimal pair on screen |
|--------------------------|--------------------------------------------------------|------------------------------------------------------|
|                          | t            | df | p    | t            | df | p    |
| Sum segmental probability | -0.14        | 45.81 | 0.89 | 0.41        | 44.98 | 0.69 |
| Sum biphone probability   | 0.62         | 41.68 | 0.54 | 0.71       | 44.09 | 0.48 |
| Neighborhood density      | 0.02         | 45.28 | 0.98 | -0.49      | 43.47 | 0.63 |
| Neighborhood frequency    | -0.71        | 33.76 | 0.49 | -1.45      | 23.35 | 0.16 |
| Sum segmental probability | -0.17        | 45.82 | 0.87 | 0.46       | 45.58 | 0.64 |
| Sum biphone probability   | 0.61         | 41.46 | 0.54 | -0.32      | 41.02 | 0.75 |
| Neighborhood density of minimal pair | -0.78 | 44.05 | 0.44 | -1.08      | 45.87 | 0.29 |
| Neighborhood frequency of minimal pair | -1.35 | 37.60 | 0.19 | 0.49       | 39.77 | 0.63 |

Fig. 7. A: Example stimulus display from the “minimal pair in lexicon”, “minimal pair on screen” condition. B: The corresponding display in the speaker’s pdf.

increase over the number (36) included in Baese-Berk & Goldrick (2009: Experiment 2). Each target item in the “minimal pair in lexicon” condition was matched for initial consonant and vowel with a target item in the “no minimal pair in lexicon” condition. Target items were balanced across levels of both independent variables for a number of phonotactic probability and phonological neighborhood measures (Vitevitch & Luce, 2004), summarized below:

- **Sum segmental probability**: log-transformed probability of each phoneme appearing in that position in the word, summed over all phonemes in the word
- **Sum biphone probability**: log-transformed probability of each sequence of two adjacent phonemes (biphone) appearing in that position in the word, summed over all biphones in the word
- **Neighborhood density**: count of real words that can be created by adding, removing, or changing one phoneme in the word
- **Neighborhood frequency**: mean frequency of occurrence of all phonological neighbors

The results of Welch’s t-tests comparing each control measure between levels of the independent variables are displayed in Table 3.

24 filler items were also included, for a total of 72 items. Six filler items began with /s/ (e.g. SIP), six began with /ʃ/ (e.g. SHIP), and 12 began with a voiced stop consonant (e.g. DESK). Those beginning with voiced stop consonants were all minimal pair competitors of experimental stimuli, while the filler items beginning with /s/ and /ʃ/ were unrelated to the experimental items. Filler items included both real and pseudowords, and, like the experimental items, varied according to whether they had a minimal pair competitor in the lexicon or not and whether the minimal pair competitor was present on the screen or not.

**Procedure**

The experiment was conducted over a Zoom call hosted by the experimenter. Before the experiment, participants were instructed to join the Zoom call from a computer (not a phone) in a quiet room with a good internet connection. At the beginning of the experiment, the experimenter changed the participant’s displayed name in the Zoom call to “speaker”, and changed their own name to “experimenter”. A lab confederate was also present in the call, whose Zoom name was changed to “listener”. The participant (henceforth “speaker”) was told that the listener was another naïve participant. Everyone in the Zoom call kept their cameras off but their microphones on for the duration of the experiment.

All instructions and stimuli were presented on a slideshow using screen sharing. During each trial, three words were presented on the shared screen: the target and two competitors, as seen in Fig. 7. Before the experiment began, the speaker was sent a pdf that matched the shared screen, except that the target word for each trial was bolded and underlined. At the start of the experiment, the speaker was instructed to arrange their computer screen so that they could see both this pdf and the shared Zoom screen at the same time. Each time the experimenter advanced to the next trial, the listener cued the speaker by saying “ready”. Then, the speaker produced the target word in the phrase “type the ___ number”. This phrase was chosen in order to encourage fluent pronunciation. Words produced in isolation, i.e., not in a larger phrase, tend to have a slower “clear speech” pronunciation, which has previously been argued to obscure CH effects (Buz et al., 2016; Wedel et al., 2018). In this phrase, the target word does not occur at a major prosodic boundary, which could condition lengthening of the target word. Moreover, the speaker was instructed not to speak slowly or pause between words, but rather to speak at a quick, conversational pace. The listener then typed the number that corresponded to the word they heard (“1” = left, “2” = center, “3” = right) into the Zoom chat. If the response was correct, the experimenter played a bell sound; if it was incorrect, the experimenter played a buzz sound. The listener (a lab confederate) did not have independent access to the correct answers, and was simply instructed to participate naturally in the experiment. In order to increase motivation, participants were informed that they

Fig. 8. Correlations among lexical control measures.
would receive an extra $5 if they finished the experiment with an accuracy of 95% or higher (all participants finished above 95%, most at 100%). The complete instructions are included in Appendix B.

After five practice trials with stimuli unrelated to the experimental stimuli, participants were instructed to ask the experimenter any questions they had. Then, the experimenter began recording the Zoom call, and each stimulus was presented twice, for a total of 144 trials (96 experimental, 48 fillers). The “minimal pair on screen” condition alternated between the two presentation lists, such that each stimulus was presented once with its minimal pair as an on-screen competitor, and once without. Four pseudo-randomizations of each presentation list were created, varying in the presentation order of the stimuli and in the relative positions of the target and competitors on the screen. Each of the six possible relative screen positions were used an equal number of times in each pseudo-randomization. In each pseudo-randomization, no two consecutive targets were minimal pairs (in any phonological feature in any position in the word) or had the same initial consonant and vowel; no two consecutive trials had the same two competitors; and no three consecutive trials had the same relative screen positions. Each pseudorandomization of the first presentation list was combined with each pseudo-randomization of the second presentation list to create 16 unique combinations. Then, the presentation order of the two lists was varied to create 32 unique combinations. Each of the 24 participants thus completed the experiment with a unique stimulus presentation order. Participants were given a short break after each quarter of the experiment (three breaks total). The entire experiment lasted less than 30 min.

Data processing

Each participant in the Zoom call was individually recorded, allowing the creation of a wav file including only the speaker’s audio. Silences between trials were removed from the wav file in Praat (Boersma & Weenink, 2021) using a custom script (Lennes, 2017), and the resulting file was force-aligned at the word and segment levels using the Montreal Forced Aligner (McAuliffe et al., 2017) with a customized English dictionary including phonetic transcriptions of the pseudoword stimulus items. Next, using AutoVOT (Sonderegger & Keshet, 2012), an automated VOT measurement algorithm was trained using 216 hand measurements in Praat from the onset of the release burst to the first zero-crossing of periodic vocal fold vibration: one measurement from each following vowel height (high, mid, low) from each consonant (/p/, /t/, /k/) from each subject (3 × 3 × 24). Even in ambiguous cases (e.g., when the release burst overlapped with some periodic energy from the preceding vowel, or when there was still some noise at the onset of periodicity), release burst and onset of periodic energy were always used as the indicators of the two events of interest. Since the AutoVOT algorithm requires the segment boundaries in the input TextGrids to be wider than the acoustic boundaries, we used a set of custom Praat scripts (Chodroff, 2019) to lengthen the segment boundaries of the word-initial voiceless stop consonants that were output by the Montreal Forced Aligner. Then, the trained algorithm was used to automatically measure

---

**Table 4**

| Variable                                           | Comp.1      | Comp.2      |
|----------------------------------------------------|-------------|-------------|
| sum segmental probability                          | 0.46        | 0.15        |
| sum biphone probability                            | 0.46        | 0.13        |
| neighborhood density                               | −0.23       | 0.61        |
| neighborhood frequency                             | −0.19       | 0.03        |
| sum segmental probability of minimal pair           | 0.49        | 0.16        |
| sum biphone probability of minimal pair             | 0.45        | 0.21        |
| neighborhood density of minimal pair                | −0.2        | 0.65        |
| neighborhood frequency of minimal pair              | −0.07       | 0.51        |

---

**Fig. 9.** Scree plot of PCA results.

**Fig. 10.** Mean VOT by condition. Error bars indicate standard error of the mean.

---

M.C. Stern and J.A. Shaw
the VOT of all 2,304 experimental tokens. In order to account for the effect of speech rate on VOT (Kessinger & Blumstein, 1997), a measure of speech rate was calculated as the duration in milliseconds from onset to offset of the trial (onset of “type” to offset of “number”).

Trials were removed from analysis for the following reasons: (1) the speaker was disfluent or there was an interruption from background noise or internet connectivity issues (138 trials, 5.99%), (2) the listener responded inaccurately (6 trials, 0.26%), (3) VOT was less than 25 ms (likely a categorical error) (91 trials, 3.95%), or (4) speech rate was greater than three standard deviations from the mean (21 trials, 0.91%). In total, 2,048 trials (89.89%) were retained for analysis.

Experiment results

Before presenting the main results of the experiment, analyzed using Bayesian mixed effects regression models, we first present an analysis of the control variables that enter into the model.

Principal component analysis of phonotactic probability and phonological neighborhood measures

Although target items were balanced across the four experimental conditions for phonotactic probability and phonological neighborhood measures (as seen in Table 3), it is still useful to include these measures as control predictors in the mixed effects regression. However, as seen in Fig. 8, there are a number of strong correlations between these control measures. These correlations could lead to a multicollinearity issue in the regression. In order to address this issue, we conducted principal component analysis (PCA) to identify the directions of maximal variability in the eight-dimensional space defined by these variables. The scree plot in Fig. 9 reveals an elbow at component 3, suggesting retention of the first two components. These two components account for 64% of the variance in the original dataset.

The loadings of the eight original variables on the two retained components—displayed in Table 4—suggest the following interpretation: component 1 generally indexes phonotactic probability (sum segmental probability and sum biphone probability of the target and minimal pair), while component 2 generally indexes phonological neighborhood density of the target and minimal pair. Importantly for the linear mixed effects regression, these two components are orthogonal to each other, addressing the issue of multicollinearity.

Fig. 9. Mean VOT target (z-axis) by voiceless target input amplitude $a_{\text{target}}$ (x-axis) and voiced competitor input amplitude $a_{\text{comp}}$ (y-axis). The center of the target distribution $p_{\text{target}} = 70$ ms is shown with a flat plane; values above this plane (yellow) indicate CH, and values below this plane (blue) indicate a trace effect. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Bayesian mixed effects model of VOT

To analyze the effect of experimental condition on VOT, we performed Bayesian mixed effects regression using the brms package (Bürkner, 2017) in R (R Core Team, 2021). The model included the following control predictors: components 1 and 2 from the PCA, speech rate, trial number, and place of articulation. Place of articulation was treatment coded with “labial” as the reference level. The other measures were scaled and centered. The experimental fixed factors were minimal pair existence, minimal pair salience, and their interaction. Both experimental fixed factors were treatment coded with “false” as the reference level. The model included random slopes by subject for all experimental fixed factors, and a random slope by item for minimal pair salience.

We followed recommendations from Franke & Roettger (2019) for fitting Bayesian mixed models. We used the default priors of the brms package: a Student’s t-distribution ($\nu = 3, \mu = 69, \sigma = 20.8$) for the intercept, a Student’s t-distribution ($\nu = 3, \mu = 0, \sigma = 20.8$) for the standard deviation of the likelihood function and the random effects, and unbiased (“flat”) priors for regression coefficients. We ran four sampling chains for 2000 iterations with a warm-up period of 1000 iterations for each model. All R-hat values (a diagnostic for convergence) were 1.0, indicating that the chains mixed successfully. Next, we removed observations with residuals to the model fit greater than 3 standard deviations from the mean (17 observations, 0.83%), and re-ran the model. Below we report the expected values of each regression coefficient under the posterior distribution and their 95% credible intervals (CrI). We consider it compelling evidence that a fixed factor reliably influenced VOT if the 95% CrI of the posterior distribution of the factor’s coefficient does not overlap with 0. Since the theoretical
Increased by a minimal pair competitor on the screen, but only when there was no reliable main effect of either minimal pair existence (labial consonants (/p/). Turning to the experimental predictors, there were quite small. Finally, place of articulation reliably affected VOT, such that VOT was longer at slower speech rates (contextually salient and 10 ms when it was (Baese-Berk & Goldrick, 2009). The reduced scope and magnitude of the pseudoword CH effect is consistent with reduced influence from the minimal pair competitor on pronunciation planning, due to reduced activation from a lexical level of planning.

**General discussion**

Cognitive mechanisms underlying speech production have been increasingly informed by the details of speech pronunciation. We proposed that a specific type of pronunciation variation, known as contrastive hyperarticulation (CH), is rooted in inhibitory neural connections between competitor lexical representations and processes of pronunciation planning. A computational model of this proposal, developed within the framework of Dynamic Field Theory (DFT), derived the results of past studies and made new predictions. Specifically, varying competitor strength—i.e., the amplitude, $\alpha$, of the competitor lexical item’s input to a dynamic neural field (DNF) governing pronunciation planning—derived the effect size in our experiment as well as the conditions in Baese-Berk & Goldrick (2009: Experiment 2). With competitor input amplitude at $-1.5$, we derived our results. Increasing the amplitude of inhibitory input to $-3$ derived the effect found for real words in the “no context” condition of Baese-Berk & Goldrick (2009:Experiment 2). Increasing inhibitory input amplitude further to $-6$ derived the effect found for real words in the “context” condition of Baese-Berk & Goldrick (2009:Experiment 2). Thus, the variety of observed CH effect magnitudes can be derived by scaling a single model parameter, competitor strength. In addition, positive (excitatory) values of this parameter derive hypoarticulation or “trace” effects of varying magnitudes, as observed in speech errors.

**Table A1**

Experimental stimuli from presentation list 1. In list 2, the “minimal pair on screen” and “no minimal pair on screen” conditions are switched. Minimal pairs differing in word-initial consonant voicing are included in parentheses, but were not presented as target items except during filler trials.

| Minimal pair on lexicon | No minimal pair on lexicon |
|-------------------------|---------------------------|
| (Labial)                |                           |
| 1. peam                 | 1. kee                    |
| (teth)                  | (geet)                    |
| 2. peck                 | 1. tep                    |
| (toke)                  | (dep)                     |
| (belch)                 | 1. keft                   |
| (dodge)                 | (godge)                   |
| 3. pid                  | 2. tob                    |
| (tsek)                  | (dode)                    |
| 4. potch                | 2. coge                   |
| (tive)                  | (gode)                    |
| (bocht)                 |                           |

| No minimal pair on lexicon (list 1) |
|-------------------------------------|
| Labial                |                         |
| 1. pag                 | 1. reg                   |
| (bag)                  | (deaf)                   |
| 2. pabble              | 2. tid                   |
| (bauble)               | (did)                    |
| 3. pathe               | 3. kide                  |
| (barhe)                | (guide)                  |
| 4. bottle              | 4. tope                  |
| (bottle)               | (dope)                   |

| Minimal pair in lexicon | No minimal pair in lexicon |
|-------------------------|---------------------------|
| Labial                  |                           |
| 1. sack                 | 1. sap                    |
| (shack)                 | (shap)                    |
| 2. sip                  | 2. sick                   |
| (ship)                  | (shick)                   |
| 3. same                 | 3. safe                   |
| (shame)                 | (shafe)                   |
| 4. sin                  | 4. sing                   |
| (shin)                  | (shing)                   |
| 5. save                 | 5. saint                  |
| (shave)                 | (shaint)                  |
| 6. sure                 | 6. soap                   |
| (shore)                 | (shoap)                   |

**Table A2**

Filler stimuli beginning with sibilants. Stimuli beginning with both /s/ and /ʃ/ were presented.

| Minimal pair in lexicon | No minimal pair in lexicon |
|-------------------------|---------------------------|
| Labial                  |                           |
| 1. pag                  | 1. reg                   |
| (bag)                   | (deaf)                   |
| 2. pabble               | 2. tid                   |
| (bauble)                | (did)                    |
| 3. pathe                | 3. kide                  |
| (barhe)                 | (guide)                  |
| 4. bottle               | 4. tope                  |
| (bottle)                | (dope)                   |

Prediction is for a small effect, even a reliable effect that is only a few milliseconds in magnitude is important. First, we address the effects of the control predictors. Component 1 (phonotactic probability) decreased VOT, such that more phonotactically probable stimuli were produced with shorter VOT ($\beta = -1.16$, 95% CrI $= [-1.88, -0.42]$). Although reliable (the 95% CrI does not overlap with 0), the effect was small: an increase in component 1 by one standard deviation is predicted to decrease VOT by only 1.16 ms. The 95% CrI for the effect of component 2 (neighborhood density) overlapped with 0 ($\beta = 0.43$, 95% CrI $= [-0.61, 1.51]$), and so was not reliable. Speech rate reliably increased VOT, such that VOT was longer at slower speech rates ($\beta = 7.07$, 95% CrI $= [6.34, 7.81]$). Trial number reliably decreased VOT: as the experiment continued, participants tended to produce shorter VOT ($\beta = -1.20$, 95% CrI $= [-1.83, -0.57]$); the magnitude of this effect was also quite small. Finally, place of articulation reliably affected VOT, such that alveolar ($\beta = 1.34$, 95% CrI $= [10.24, 16.96]$) and velar ($\beta = 12.60$, 95% CrI $= [9.19, 15.89]$) consonants had longer VOT than labial consonants ($\beta = 0.12$). Turning to the experimental predictors, there was no reliable main effect of either minimal pair existence ($\beta = 0.04$, 95% CrI $= [-3.25, 3.39]$) or minimal pair salience ($\beta = -0.29$, 95% CrI $= [-2.10, 1.49]$). However, there was a positive interaction: VOT was increased by a minimal pair competitor on the screen, but only when that minimal pair competitor was also a real word ($\beta = 2.50$, 95% CrI $= [0.04, 4.93]$). As predicted, the magnitude of this CH effect was small compared with the previously observed effect on real words: 2.5 ms. The interaction is visualized in Fig. 10.

**Discussion**

The experiment results are consistent with a basis for CH in the real-time planning and production of speech, since pseudowords have no long-term lexical representations. Importantly, the CH effect in pseudowords was smaller in scope (only appearing when there was a real word minimal pair that was also present as an on-screen competitor) and magnitude (model estimate $= 2.5$ ms) compared to the CH effects observed with real words, i.e., 5 ms when the minimal pair was not contextually salient and 10 ms when it was (Baese-Berk & Goldrick, 2009). The reduced scope and magnitude of the pseudoword CH effect is consistent with reduced influence from the minimal pair competitor on pronunciation planning, due to reduced activation from a lexical level of planning.

--

5 It is worth highlighting the fact that we model inhibition as coming from lexical representations, rather than sublexical representations like phonemes or syllables. The reason for this is empirical: CH (as well as other forms of context-based phonetic enhancement and reduction; Hall et al., 2018) appear to operate primarily on lexical items, rather than sublexical units. Both the /p/ in PET and the /ʃ/ in PEP have a sublexical competitor /b/; CH is observed in the pronunciation of PET relative to PEP because of the presence of a lexical competitor in the former (BET) but not the latter (“BEP”) case.
Table A3
Unrandomized list of trials including target items and both competitors for practice trials and both presentation lists.

| list    | trial type | minimal pair in lexicon | minimal pair on screen | place of articulation | target | competitor 1 | competitor 2 |
|---------|------------|-------------------------|------------------------|-----------------------|--------|--------------|--------------|
| practice | practice   | —                       | —                      | —                     | vent   | vine         | chair        |
| practice | practice   | —                       | —                      | —                     | teeef  | teeth        | phone        |
| practice | practice   | —                       | —                      | —                     | pite   | pote         | thick        |
| practice | practice   | —                       | —                      | —                     | beef   | beeth        | door         |
| 1        | experimental | yes                    | yes                    | labial                | peam   | beam         | sack         |
| 1        | experimental | yes                    | yes                    | labial                | pelch  | belch        | sip          |
| 1        | experimental | yes                    | yes                    | labial                | pid    | bid          | same         |
| 1        | experimental | yes                    | yes                    | labial                | poch   | botch        | shap         |
| 1        | experimental | yes                    | yes                    | alveolar              | teth   | death        | shick        |
| 1        | experimental | yes                    | yes                    | alveolar              | tojde  | dodge        | shafe        |
| 1        | experimental | yes                    | yes                    | alveolar              | tejk   | desk         | sin          |
| 1        | experimental | yes                    | yes                    | alveolar              | tive   | dive         | save         |
| 1        | experimental | yes                    | yes                    | velar                 | keese  | geese        | sore         |
| 1        | experimental | yes                    | yes                    | velar                 | cosh   | gosh         | sing         |
| 1        | experimental | yes                    | yes                    | velar                 | kig    | gij          | saint        |
| 1        | experimental | yes                    | yes                    | velar                 | kulp   | gulp         | soap         |
| 1        | experimental | yes                    | no                     | labial                | pag    | sack         | shack        |
| 1        | experimental | yes                    | no                     | labial                | pabble | sip          | ship         |
| 1        | experimental | yes                    | no                     | labial                | pale   | same         | shame        |
| 1        | experimental | yes                    | no                     | labial                | pottle | sin          | shin         |
| 1        | experimental | yes                    | no                     | alveolar              | teff   | save         | shafe        |
| 1        | experimental | yes                    | no                     | alveolar              | tid    | sore         | shore        |
| 1        | experimental | yes                    | no                     | alveolar              | tupe   | sap          | shap         |
| 1        | experimental | yes                    | no                     | alveolar              | tope   | sick         | shick        |
| 1        | experimental | yes                    | no                     | velar                 | tof    | deld         | shaf         |
| 1        | experimental | yes                    | no                     | velar                 | kiv    | sing         | shing        |
| 1        | experimental | yes                    | no                     | velar                 | kide   | saint        | shaint       |
| 1        | experimental | yes                    | no                     | velar                 | kest   | soap         | shoap        |
| 1        | experimental | yes                    | yes                    | velar                 | kip    | gip          | saint        |
| 1        | experimental | no                     | yes                    | velar                 | kulik  | gulk         | soap         |
| 1        | experimental | no                     | yes                    | velar                 | paz    | sack         | shack        |
| 1        | experimental | no                     | yes                    | velar                 | pakkle | sip          | ship         |
| 1        | experimental | no                     | yes                    | velar                 | pame   | same         | shae         |
| 1        | experimental | no                     | yes                    | velar                 | ponle  | sin          | shin         |
| 1        | experimental | no                     | no                     | alveolar              | tged   | save         | shave        |
| 1        | experimental | no                     | no                     | alveolar              | tive   | sore         | shore        |
| 1        | experimental | no                     | no                     | alveolar              | toog   | sap          | shap         |
| 1        | experimental | no                     | no                     | alveolar              | tobe   | sick         | shick        |
| 1        | experimental | no                     | no                     | velar                 | koom   | safe         | shafe        |
| 1        | experimental | no                     | no                     | velar                 | kide   | sing         | shing        |
| 1        | experimental | no                     | no                     | velar                 | kife   | saint        | shaint       |
| 1        | experimental | no                     | no                     | velar                 | lelf   | soap         | shoap        |
| 1        | filler      | yes                    | yes                    | J                     | shack  | sack         | pag          |
| 1        | filler      | yes                    | yes                    | j                     | ship   | sip          | pbble        |
| 1        | filler      | yes                    | yes                    | j                     | shame  | same         | path        |
| 1        | filler      | yes                    | no                     | j                     | shin   | tesk         | desk         |
| 1        | filler      | yes                    | no                     | j                     | shave  | tive         | dive         |
| 1        | filler      | yes                    | no                     | j                     | shore  | keese        | geese        |
| 1        | filler      | no                     | yes                    | s                     | sap    | shap         | tope         |
| 1        | filler      | no                     | yes                    | s                     | sick   | shick        | tope         |
| 1        | filler      | no                     | yes                    | s                     | safe   | shafe        | coof         |
| 1        | filler      | no                     | no                     | s                     | sing   | codge        | godge        |
| 1        | filler      | no                     | no                     | s                     | saint  | kip          | gip          |
| 1        | filler      | no                     | no                     | s                     | soap   | kulp         | gulp         |
| 1        | filler      | yes                    | yes                    | labial                | beam   | peam         | sack         |
| 1        | filler      | yes                    | yes                    | alveolar              | dodge  | tojde        | shafe        |
| 1        | filler      | yes                    | yes                    | velar                 | gig    | kig          | saint        |
| 1        | filler      | yes                    | no                     | labial                | belch  | same         | shae         |
| 1        | filler      | yes                    | no                     | alveolar              | desk   | sore         | shor         |
| 1        | filler      | yes                    | no                     | velar                 | geese  | saint        | shaint       |
| 1        | filler      | no                     | yes                    | labial                | beeb   | peeb         | shack        |
| 1        | filler      | no                     | yes                    | alveolar              | deld   | teld         | shin         |
| 1        | filler      | no                     | yes                    | velar                 | gosh   | cosh         | sing         |

(continued on next page)
| list  | trial type | minimal pair in lexicon | minimal pair on screen | place of articulation | target | competitor 1 | competitor 2 |
|-------|------------|-------------------------|------------------------|-----------------------|--------|--------------|--------------|
| 1     | filler     | no                      | no                     | labial                | hodge  | sip           | ship         |
| 1     | filler     | no                      | no                     | alveolar              | dibe   | sap           | shap         |
| 1     | filler     | no                      | no                     | velar                 | geet   | sing          | shing        |
| 2     | experimental | yes                   | no                     | labial                | peam   | sack          | shack        |
| 2     | experimental | yes                   | no                     | labial                | pelch  | sip           | ship         |
| 2     | experimental | yes                   | no                     | labial                | pid    | same          | shame        |
| 2     | experimental | yes                   | no                     | labial                | potch  | sin           | shin         |
| 2     | experimental | yes                   | no                     | alveolar              | teth   | save          | shave        |
| 2     | experimental | yes                   | no                     | alveolar              | toge   | sore          | shore        |
| 2     | experimental | yes                   | no                     | alveolar              | teff   | deaf          | shick        |
| 2     | experimental | yes                   | no                     | alveolar              | tid    | did           | shafe        |
| 2     | experimental | yes                   | no                     | alveolar              | tupe   | dupe          | sin          |
| 2     | experimental | yes                   | no                     | alveolar              | tope   | dope          | save         |
| 2     | experimental | yes                   | no                     | velar                 | coof   | goof          | sore         |
| 2     | experimental | yes                   | no                     | velar                 | kiv    | give          | sing         |
| 2     | experimental | yes                   | no                     | velar                 | kilde  | guide         | saint        |
| 2     | experimental | yes                   | no                     | velar                 | kest   | guest         | soap         |
| 2     | experimental | no                    | no                     | labial                | peeb   | sack          | shack        |
| 2     | experimental | no                    | no                     | labial                | peft   | sip           | ship         |
| 2     | experimental | no                    | no                     | labial                | pin    | same          | shame        |
| 2     | experimental | no                    | no                     | labial                | podge  | sin           | shin         |
| 2     | experimental | no                    | no                     | alveolar              | tep    | save          | shave        |
| 2     | experimental | no                    | no                     | alveolar              | tob    | sore          | shore        |
| 2     | experimental | no                    | no                     | alveolar              | teled  | sap           | shap         |
| 2     | experimental | no                    | no                     | alveolar              | tibe   | sick          | shick        |
| 2     | experimental | no                    | no                     | velar                 | keet   | safe          | shafe        |
| 2     | experimental | no                    | no                     | velar                 | codge  | sing          | shing        |
| 2     | experimental | no                    | no                     | velar                 | kip    | saint         | shaint       |
| 2     | experimental | no                    | no                     | velar                 | zahl   | soap          | shope        |
| 2     | experimental | no                    | no                     | velar                 | kide   | gidge         | sing         |
| 2     | experimental | no                    | no                     | velar                 | kile   | gife          | saint        |
| 2     | experimental | no                    | yes                    | velar                 | keft   | gife          | soap         |
| 2     | experimental | no                    | no                     | velar                 | j      | shack         | pag          |
| 2     | filler      | yes                    | no                     | velar                 | j      | ship          | pabble       |
| 2     | filler      | yes                    | no                     | velar                 | j      | shame         | pathe        |
| 2     | filler      | yes                    | yes                    | velar                 | j      | shap          | sap          |
| 2     | filler      | yes                    | yes                    | velar                 | j      | shick         | sick         |
| 2     | filler      | yes                    | yes                    | velar                 | j      | shafe         | safe         |
| 2     | filler      | no                     | no                     | velar                 | s      | sin           | tupe         |
| 2     | filler      | no                     | no                     | velar                 | s      | save          | tope         |
| 2     | filler      | no                     | no                     | velar                 | s      | sore          | coof         |
| 2     | filler      | yes                    | yes                    | velar                 | s      | sing          | shing        |
| 2     | filler      | yes                    | yes                    | velar                 | s      | saint         | shaint       |
| 2     | filler      | yes                    | no                     | labial                | s      | soap          | shope        |
| 2     | filler      | yes                    | no                     | labial                | s      | beam          | sack         |
| 2     | filler      | yes                    | no                     | labial                | s      | alveolar      | dodge        |
| 2     | filler      | yes                    | no                     | velar                 | s      | gig           | saint        |
| 2     | filler      | yes                    | no                     | labial                | s      | alveolar      | bottle       |
| 2     | filler      | yes                    | no                     | velar                 | s      | did           | shade        |
| 2     | filler      | yes                    | no                     | velar                 | s      | velar         | give         |
| 2     | filler      | no                     | no                     | labial                | s      | beeb          | sack         |
| 2     | filler      | no                     | no                     | alveolar              | s      | deid          | sin          |
| 2     | filler      | no                     | no                     | velar                 | s      | gosh          | sing         |
| 2     | filler      | no                     | yes                    | labial                | s      | baz           | paw          |
| 2     | filler      | no                     | yes                    | velar                 | s      | deg           | teg          |
| 2     | filler      | no                     | yes                    | velar                 | s      | left          | keft         |

M.C. Stern and J.A. Shaw

Journal of Memory and Language 132 (2023) 104443
Our theory thus offers a unified explanation of these apparently disparate phenomena. The model also makes predictions regarding response time, specifically: (i) a positive correlation between competitor strength and response time, and (ii) a positive correlation between response time and the magnitude of CH. These predictions remain to be tested. Broadly, the explicit incorporation of both temporal and feature gradiences into a single model of pronunciation planning allows the generation of a rich and precise set of empirical predictions (Roon & Gafos, 2016). For instance, in the model, activation peaks narrow over time via the combination of lateral excitation and lateral inhibition. This predicts a relationship between response time and response variability, such that—when paper else—shorter response times should correspond with more variable responses, because the activation peak corresponding to the planned response has less time to narrow. Such a relationship between response time, width of neural activation peak, and response variability has been observed in manual reaching movements (Erlhagen & Schoner, 2002) and theus monkey motor cortex (Georgopoulos et al., 1986)—it remains to be tested in human speech.

A reviewer raised the possibility that it is the relationship between activation of the target word and activation of the minimal pair competitor—that determines the pronunciation outcome. In order to investigate this issue, we ran simulations using the same parameters described in Section ‘Dynamic neural field model of VOT planning’, but varying both competitor and target input amplitude. We varied the competitor input amplitude, \( a_{\text{comp}} \), from -6 to 5 and the target input amplitude, \( a_{\text{target}} \), from 5 to 10, both in steps of 0.5. We simulated 500 productions in each condition and recorded each VOT target. Fig. 11 displays the results. Consistent with the results in Section ‘Simulation results’, the primary predictor of the qualitative behavior of the system is the competitor input amplitude, \( a_{\text{comp}} \). When \( a_{\text{comp}} \) is negative, CH is observed; when \( a_{\text{comp}} \) is positive, a trace effect is observed. The magnitude of each type of effect correlates with the absolute value of \( a_{\text{comp}} \). In contrast, the target input amplitude, \( a_{\text{target}} \), modulates the magnitude of competitor influence. Greater \( a_{\text{target}} \) corresponds with less competitor influence, i.e., smaller CH when competitor input is inhibitory, and a smaller trace effect when CH is excitable. Interestingly, there is an asymmetry between CH and trace effects in this regard. The magnitude of CH is relatively robust to changes in \( a_{\text{target}} \): when competitor input is strongly inhibitory (\( a_{\text{comp}} = -6 \)), mean CH ranges from 6.1 ms (\( a_{\text{target}} = 10 \)) to 10.4 ms (\( a_{\text{target}} = 5 \)). However, when competitor input is strongly excitative (\( a_{\text{comp}} = 5 \)), the magnitude of the trace effect ranges from 8.4 ms (\( a_{\text{target}} = 10 \)) to 24.9 ms (\( a_{\text{target}} = 5 \)). This is because as \( a_{\text{comp}} \) approaches \( a_{\text{target}} \), the two hills of activation begin to merge into a single peak, approximately equidistant from the centers of the two input distributions \( p_{\text{comp}} \) and \( p_{\text{target}} \). No such mechanism operates when \( a_{\text{comp}} \) is negative. Thus, while these post-hoc simulations support our claim that the magnitude of CH is primarily determined by activation of a minimal pair competitor, it also suggests the need for follow-up work investigating the influences of target and competitor activation in relation to CH and trace effects.

In the model presented here, CH is derived from a mechanism internal to the speech production system, i.e., inhibitory input from competitor lexical representations to pronunciation planning fields. The listener has no explicit role in the model. However, the speaker-internal inhibitory mechanism is compatible with audience design theories of CH (Buz et al., 2016; Munson & Solomon, 2004; Schertz, 2013; Wright, 2004). Inhibition can be seen to implement audience design on the part of the speaker, since inhibition has the effect of making pronunciations more easily perceivable for the listener. In this way, a potential theoretical dichotomy between “speaker-internal” vs. “audience design” mechanisms of CH can be resolved. The mechanism itself is speaker-internal, but it may serve a listener-oriented function. This conceptualization has the potential to shed light on previously observed listener-driven modulation of CH effects. For instance, Buz et al. (2016) observed an increase in the magnitude of CH following explicit misrecognition by a simulated listener. In the present framework, there are at least two potential mechanisms that might drive this effect. First, it is possible that as speakers sense an increased risk of listener misrecognition, this causes an increase in the activation of competitors, increasing the magnitude of inhibitory input from competitors to pronunciation planning fields, thus increasing the magnitude of CH. This mechanism would capture the intuition that strongly intending not to say a word actually increases mental activation of that word, and is consistent with the observation in Buz et al. (2016) that listener-driven modulation of CH was only observed on trials where the minimal pair competitor was present. Another possibility is that increased speaker awareness of the risk of listener misrecognition modulates the function relating lexical activation to field input, leading to greater inhibitory input given the same magnitude of competitor activation (see further discussion of this function below). Further modeling and empirical work would be necessary to refine and distinguish between these proposed mechanisms.

The difference in CH magnitude between real words and pseudo-words supports a role for lexical representations in contributing to CH. So far, we have considered the hypothesis that this contribution takes place primarily during the planning of individual utterances, increasing the strength of inhibitory input from competitor lexical items to pronunciation planning fields. However, another possibility, discussed in Section ‘Possible sources of constraint hyperarticulation’, is that lexical representations contribute to CH on a slower timescale, via updating of detailed episodic memories of perceived and produced pronunciations (e.g., Goldinger, 1998; Pierrehumbert, 2001, 2002). In the present framework, this amounts to lexical representations updating their input distributions, parameterized as \( p \) and \( w \). It is possible that the coupling relationships between lexical representations and pronunciation planning fields are somewhat idiosyncratic, such that each lexical representation projects a unique set of inputs to pronunciation planning fields. These coupling relationships might change subtly over the lifetime of language users based on experience (Gafos & Kirov, 2009), leading to the observed idiosyncratic phonetic differences between lexical items (e.g., Tang & Shaw, 2021). It is possible that lexical representations contribute to CH via mechanisms on both timescales: interactive activation during speech planning, and long-term shifts in distributions of episodic memories. A combination of real-time and long-term effects on pronunciation details has previously been observed in the relationship between contextual predictability and phonetic cues of prominence like duration, pitch, and intensity (Seyfarth, 2014; Sökşuthy & Hay, 2017; Tang & Shaw, 2021). The contribution of a long-term mechanism is possibly apparent in the fact that VOT was overall slightly higher in the real words from Baese-Berk & Goldrick (2009) (baseline condition = 72 ms) compared to the pseudowords from the present experiment (baseline = 69 ms). However, the present experiment was not an exact replication of Baese-Berk & Goldrick (2009); most notably, our experiment was conducted over Zoom rather than in person. This may have affected the degree of clear speech employed by participants, which has been argued to affect the magnitude of CH effects (Wedel et al., 2018). Thus, it is difficult to interpret any direct quantitative comparison between the results of our study and those of Baese-Berk & Goldrick (2009).

Deriving the range of observed CH and trace effects from competitor strength raises a number of novel research questions. In our simulations, we varied competitor strength systematically to observe its effects on pronunciation. Presumably, the ranges of competitor strength derive from other factors that can be built into a more elaborate model. In

---

6 VOT targets reported in the simulation include some cases in which the field did not stabilize within the timeframe of the simulation, which is possible given our method of interpreting field dynamics in terms of VOT targets (see footnote 5).
future work, we plan to derive competitor strength and polarity directly from the lexicon. This will enable us to define the magnitude of a as a function of lexical activation, such that more active lexical items project stronger input to pronunciation planning DNFs. We also posit that the polarity of a (i.e., excitatory vs. inhibitory) varies, deriving the difference between CH (from inhibitory competitor input) and trace effects (from excitatory competitor input). What determines the polarity of a? One possibility is that input polarity is a function of the similarity of the current state of the field to the lexical representation’s preferred state of the field, i.e., the state that would be induced by excitatory input from that lexical representation. For instance, in the example in Fig. 1, since PUN becomes active before BUN, by the time BUN is active enough to affect pronunciation planning DNFs, the VOT field is in a state more consistent with a voiceless production than a voiced production. This induces the BUN-to-VOT projection to be inhibitory. These elaborations would derive competitor strength and polarity directly from the lexicon, leading to new predictions for how pronunciation varies across words.

Another future direction involves deriving speech errors. Noise in the relationship between field state and input polarity offers a possible mechanism underlying speech errors: sometimes, lexical-to-phonological projection is excitatory, even when the field state differs from the lexical item’s preferred state, leading to a trace effect. The fact that visual input to the minimal pair competitor increases the magnitude of CH (by hypothesis, by increasing the magnitude of inhibitory lexical-to-phonological input) supports the notion that the magnitude of input is a positive monotonic function of lexical activation, regardless of the polarity of input. Noise around the polarity of the input could cause a strong inhibitory input to flip to a strong excitatory input, driving the field towards a state consistent with the pronunciation of a lexical competitor, i.e., a speech error.

In our account, CH derives from decreased activation in certain regions of a pronunciation planning field caused by inhibitory input from a minimal pair competitor. A reviewer raises the possibility that this metrically specific decrease in neural activation might be caused by a different mechanism, like habituation via synaptic fatigue. Pursuing this possibility would raise a number of interesting questions. For instance, why would competitor activation cause synaptic fatigue in fields governing dimensions of pronunciation on which the competitor differs from the target word, but not in fields governing dimensions on which the competitor and target overlap? Moreover, how do patterns of synaptic fatigue differ between non-errorful and errorful speech, such that CH occurs in the former but trace effects occur in the latter? We feel that, at this point, our inhibition account is the most promising in terms of coverage of existing empirical facts and generation of novel predictions, but it would be useful to carefully formulate alternatives in order to compare their predictions with those of our account.

To take a broader theoretical perspective on our proposal, we can ask why the neurocognitive mechanisms underlying speech planning would have evolved to incorporate both excitatory and inhibitory interactions between lexical and phonological levels of planning. In other words, why would activating one lexical representation affect the activation of other similar lexical representations (competitors), and why would competitors affect pronunciation planning? From the perspective of production, interactive lexical-phonological activation can be seen to have a facilitative effect, speeding up the process of activation peak stabilization in pronunciation planning fields via additional input from competitors. Thus, having competitors facilitates pronunciation. Of course, activating lexical competitors may allow them to influence pronunciation planning introduces a risk that the planning process might converge on a competitor pronunciation, rather than the target. In fact, the cost of excitatory interaction is likely often seen in speech errors, when a competitor is produced instead of the target. Excitatory projections bring the benefit of faster speech production and the risk of occasional errors. Inhibitory lexical-to-phonological projection can be seen to mitigate this risk. By projecting inhibitory input to just those pronunciation planning fields that differentiate the competitor from the target (via the mechanism described above), competitors are able to facilitate planning of dimensions of pronunciation which are shared with the target without disrupting planning of dimensions which are not shared with target. Moreover, as discussed above, this inhibitory mechanism serves the additional function of making pronunciations more easily perceivable by listeners. Both excitatory and inhibitory coupling mechanisms between the lexicon and pronunciation planning fields have clear roles in a speech production system evolved for efficiency.

Explicitly incorporating these hypothesized lexical-phonological coupling mechanisms into the present model has the potential to shed new light on the mechanisms underlying the effects of phonological neighborhoods on speech production. Different studies have identified facilitation (Vitevitch, 2002), inhibition (Gordon & Kurczek, 2014), hyperarticulation (Munson & Solomon, 2004), and hyparticulation (Gahl et al., 2012) induced by phonological neighborhood density. Perhaps this diversity of findings is a result of the coarseness of existing phonological neighborhood density measures: most commonly, the number of real words that can be created by adding, removing, or changing a single phoneme. In the framework presented here, neighborhood density itself is not expected to have a consistent effect on pronunciation planning. Rather, different types of neighbors are expected to have different effects on different dimensions of pronunciation. In particular, neighbors are predicted to inhibit dimensions that differentiate them from the target and excite dimensions on which they overlap with the target. The aggregate of these effects from all neighbors will ultimately derive the pronunciation targets of a particular utterance, and the time it takes to plan these targets. Crucially, the simple number of neighbors should be a weak predictor of target values and target planning times. For instance, a word with many more neighbors that differ in initial consonant voicing than neighbors that share initial consonant voicing is likely to have a hyperarticulated VOT, compared to a word with more neighbors that share voicing than that differ in voicing—even if both words have the same neighborhood density. The complexity of the influences of phonological neighborhoods on speech production points to the utility of a computational modeling approach like the one pursued here. A mathematical formalization of the processes of pronunciation planning and their coupling to lexical planning allows the generation of precise quantitative predictions in cases where such predictions could not possibly be intuited based on a verbally articulated theory alone. We believe this is a promising direction for future work and may lead to new empirical predictions linking the temporal dynamics of speech planning to the phonetic details of pronunciation.

Conclusion

We investigated the real-time mechanisms contributing to contrasting hyperarticulation (CH) during the planning of individual utterances. We demonstrated—using a model of pronunciation planning based on Dynamic Field Theory (DFT)—that CH is derivable from inhibitory projections from the minimal pair lexical representation to phonological levels of planning. We also showed that the magnitude of inhibitory projection correlates with the magnitude of CH, and that the magnitude of excitatory projection correlates with the magnitude of the phonetic trace effect in speech errors. We thus derived the observed range of CH and trace effect magnitudes by scaling a single model parameter, competitor strength, offering a unified explanation of an apparently disparate set of phenomena. We tested some predictions of the model with a novel experiment, very similar to Experiment 2 in Baese-Berk & Goldrick (2009) but with pseudoword stimuli. Pseudowords demonstrated CH effects, consistent with a contribution of real-time mechanisms to CH. However, the scope and magnitude of CH in pseudowords was reduced compared to CH in real words, consistent with a role for interactive activation between lexical and phonological representations. We outlined directions for future work which are suggested by the present study.
Appendix A. Stimuli

See Table A1, Table A2 and Table A3.

Appendix B. Instructions

Welcome to the experiment! Please leave your camera off but your microphone on (unmutted). The experimenter will randomly change one of your Zoom names to “speaker”, and the other to “listener”. “Speaker”: you will receive a pdf in the Zoom chat. Please open this pdf to the first page, and arrange your computer screen so that you can see both the pdf and the shared zoom screen at the same time.

Your goal is to communicate successfully by working together. (Please say “ok” when you’ve finished reading the slide).

Each trial in the experiment will go like this: Three words will appear on the screen. Some of the words will be real words of English, and others will be made up words.

Listener: After the words appear on the screen, say “ready”. Then, the speaker will say the target word for you. If you think the target is the word on the left side of the screen, type “1” into the Zoom chat. If you think it is the word in the center, type “2”. If you think it is the word on the right, type “3”.

Listener: It is your task to listen closely to make sure you select the correct target word. During the experiment, please do not say anything except “ready”.

Speaker: In the pdf you received, the target word for each trial is bolded and underlined. Before each trial, make sure the three words on the pdf match the three words on the shared zoom screen.

Speaker: After the listener says “ready”, say the target word in the phrase “type the_number”, for example “type the sample number”. If the target is not a real word of English, just pronounce it in the way that seems most natural. It is your task to make sure the listener chooses the correct word.

Speaker: Please do not speak slowly or pause between words. Rather, speak at a fluent, conversational pace. During the experiment, please do not say anything except the target word in the phrase “type the_number”.

After each trial, you will hear a bell sound if you are correct, or a buzz sound if you are incorrect. If you finish the experiment with an overall accuracy of 95% or more, you will each receive a bonus $5. Let’s do a little practice before we start.
