



# Inhibitory modulation of speech trajectories: Evidence from a vowel-modified Stroop task

Sara D. Beach <sup>a</sup> and Caroline A. Niziolek <sup>a,b</sup>

<sup>a</sup>Waisman Center, University of Wisconsin–Madison, Madison, WI, USA; <sup>b</sup>Department of Communication Sciences and Disorders, University of Wisconsin–Madison, Madison, WI, USA

## ABSTRACT

How does cognitive inhibition influence speaking? The Stroop effect is a classic demonstration of the interference between reading and color naming. We used a novel variant of the Stroop task to measure whether this interference impacts not only the response speed, but also the acoustic properties of speech. Speakers named the color of words in three categories: congruent (e.g., *red* written in red), color-incongruent (e.g., *green* written in red), and vowel-incongruent — those with partial phonological overlap with their color (e.g., *rid* written in red, *grain* in green, and *blow* in blue). Our primary aim was to identify any effect of the distractor vowel on the acoustics of the target vowel. Participants were no slower to respond on vowel-incongruent trials, but formant trajectories tended to show a bias away from the distractor vowel, consistent with a phenomenon of *acoustic inhibition* that increases contrast between confusable alternatives.

## ARTICLE HISTORY

Received 6 September 2022  
Revised 12 January 2024  
Accepted 15 January 2024

## KEYWORDS

Speech; vowels; inhibition; competition; Stroop

## Introduction

In everyday conversation as well as in the speech laboratory, words compete for selection. Evidence of competitors can be found in subtle variations in the way that an individual pronounces a word. For example, close analysis of a tongue-twister production such as “she sells seashells” might find that the acoustic energy of /s/ partially resembles that of /ʃ/. These subthreshold modulations reflect a hybrid of the target and the competitor, suggesting the cascading activation of more than one phonological plan (Goldrick & Blumstein, 2006) and/or multiple concurrent articulatory gestures (Goldstein et al., 2007; Mowrey & MacKay, 1990). On the other hand, some forms of competition naturally arise from the linguistic system but seem to have the opposite acoustic effect: the existence of a minimal-pair neighbour (e.g., *sheep* and *ship*, but not *steep* and *\*stip*) causes speakers to increase the phonetic distance between the vowels that distinguish them (Clopper & Tamati, 2014; Wedel et al., 2018). Thus, in some contexts, the competitor intrudes upon the target, yielding an intermediate production, but in other contexts, the competitor’s influence is contrastive, exaggerating the distance between itself and the intended target.

The goal of the current study was to determine how sublexical competition from a salient distractor affects the acoustic properties of speech. To this end, we took advantage of the classic Stroop (1935) color-word effect, in which reading interferes with color naming such that individuals are slower to name an incongruent text color (e.g., the word *red* written in green) than a congruent text color (e.g., the word *blue* written in blue). The Stroop task is a powerful and reliable way to activate competing information (MacLeod, 1991). In fluent readers, visual word presentation automatically activates acoustic and phonological as well as semantic representations (Booth et al., 1999; Mousikou et al., 2015; Rayner et al., 1995). Variations of the Stroop paradigm have previously shown that phonological characteristics of the written word (henceforth “the distractor”) influence the speed of performance (Berent & Marom, 2005; Dennis & Newstead, 1981; Sumiya & Healy, 2004). As evidenced by errors and slowed reaction times, a top-down control process — referred to as *cognitive inhibition* or *intentional inhibition* (Harnishfeger, 1995; Pompon et al., 2015) — is required to overcome the distractor and articulate the target color name. To our knowledge, however,

no study has examined whether resolving Stroop-like interference has consequences for *how* — not only how quickly — the target is pronounced.

How might this interference manifest in pronunciation? One hypothesis is that the prepotent activation of the distractor spreads to the articulatory level: that is, the salient but irrelevant orthographic stimuli in this task are unconsciously integrated into speech motor plans despite attempts to ignore them. If multiple sound representations are active at the same time, the result may be a kind of phonetic average, a biasing of target acoustics towards the distractor. This “attraction” outcome is plausible under models of speech production in which coactivation of multiple competing representations cascades to downstream levels of processing (Rapp & Goldrick, 2000). Another hypothesis is that it is the ultimately successful inhibition of the distractor which spreads downstream to affect articulation. If two representations exist as neighbours in a continuous feature space, the suppression of one region (by cognitive inhibition, perhaps) may drive dissimilation of the other representation, biasing its acoustics away from those of the suppressed distractor (Stern & Shaw, 2023). This “repulsion” outcome highlights how the communicative and informational context may generate more distinct, contrastive pronunciations (Tilsen, 2013): faced with two competing representations, speakers avoid pronunciations close to the interfering word.

Here we administered the traditional congruent and color-incongruent Stroop conditions, and introduced a novel “vowel-incongruent” condition in which distractors were not conceptually-related color names but words sharing the majority of their orthographic and phonological structure with the targets. Each vowel-incongruent distractor differed from its corresponding target only in its vowel nucleus, allowing us to measure, as our primary aim, the effect of a distractor vowel on the pronunciation of another vowel that shares the same consonant frame. We therefore focused our acoustic analyses on the first and second formant frequencies (F1 and F2), which together provide the primary cues to vowel identity. By comparing how a word is pronounced with congruent vs. incongruent external cues, we aimed to determine whether responses in the Stroop framework are better characterized by attraction (of the target’s formants toward the

distractor) or by repulsion (of the target’s formants away from the distractor). We also considered alternative potential outcomes: incongruent information may disrupt speech planning in a general way, yielding more variable utterances (Dromey & Bates, 2005; Kleinow & Smith, 2006; Oostwoud Wijdenes et al., 2016; see also Goldrick & Blumstein, 2006), or it may have no effect on spoken acoustics, such that cognitive interference slows vocal responses but is otherwise independent of the implementation details of articulation. Across two participant samples, we find reliable evidence that cognitive inhibition of a phonologically-similar written word during speech production biases spoken acoustics away from that word, suggesting a phenomenon of *acoustic inhibition*.

## Materials and methods

### Participants

Two independent samples were recruited for this study. The original sample included 17 healthy adults (9 females and 8 males; mean age = 19.70 years, SD = 0.77 years) and the replication sample included 18 healthy adults (13 females and 5 males; mean age = 20.94 years, SD = 2.29 years). All were native speakers of American English and reported normal hearing, normal or corrected-to-normal vision, and no history of speech or language disorders. Participants provided informed consent and were compensated for their time. The study was overseen by the Institutional Review Boards of Boston University (original sample) and the University of Wisconsin–Madison (replication sample). Except where stated otherwise, results refer to data from the original sample.

### Stimuli

Experimental stimuli are shown in Figure 1(a). Congruent stimuli consisted of the words *red* (/ɹɛd/), *green* (/gɹiːn/), and *blue* (/blu/) printed in their corresponding colors (RGB values: red = [255, 0, 0]; green = [84, 204, 51]; blue = [0, 50, 255]). Vowel-incongruent stimuli consisted of the words *rid* (/ɹɪd/) and *rad* (/ɹæd/) printed in red; *grin* (/gɹɪn/) and *grain* (/gɹeɪn/) printed in green; and *bleed* (/bliːd/) and *blow* (/blou/) printed in blue. In other words, to the

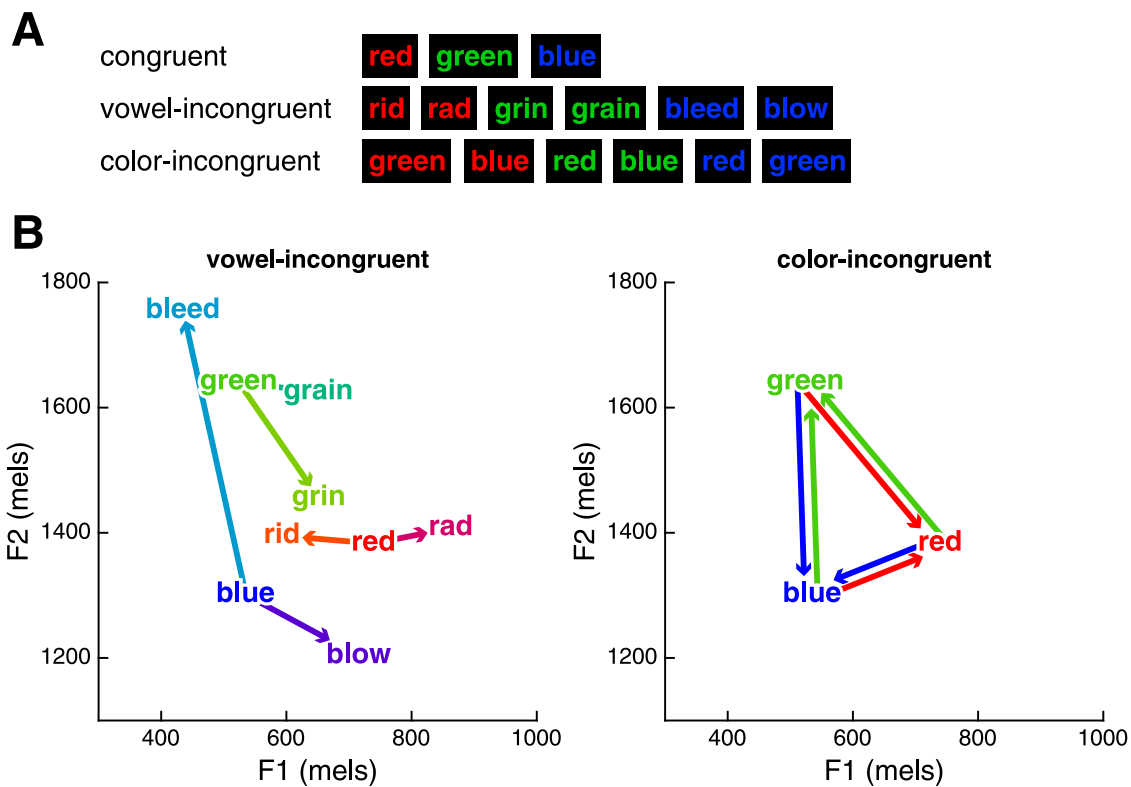


**Figure 1.** Experimental stimuli. (A) Visual stimuli shown to participants during the main experiment. (B) Word productions plotted in vowel space: the centre of each word is located at the (F1, F2) coordinate representing its median formant values, measured at the middle 50 percent of the syllable, when pronounced by participants in a neutral context prior to the main experiment. Lines between words illustrate the change in formants that would constitute attraction by the distractor. (Note that the shades of red, green, and blue in (B, left) did not appear in the experiment, but introduce the color-coding scheme for subsequent figures in which results for each distractor have a characteristic color.)

extent possible in the English lexicon, the consonant frame of the target color name was held constant and the vowel was replaced with one that would yield a familiar, pronounceable word. Color-incongruent stimuli consisted of the words *green* and *blue* printed in red; *red* and *blue* printed in green; and *red* and *green* printed in blue. All words were presented in lowercase Arial typeface and centred on a black background.

### Procedure

Participants performed a modified version of the Stroop task lasting between 30 and 45 min. They were seated in front of a computer screen in a quiet testing room and spoke into an adjustable, head-worn condenser microphone (AKG Pro Audio C520) placed approximately one inch from the corner of the mouth. Verbal responses were recorded at a sampling rate of 11025 Hz. In all cases, stimuli were presented one at a time for 1000 ms and followed

by a blank screen for a randomly selected duration of between 750 and 1500 ms. The task was administered using MATLAB (Mathworks, Natick, MA) with Psychtoolbox extensions (Brainard, 1997).

We first established, in a neutral context, formant frequencies for the nine words in the experiment. Each was presented in white text ten times in a random order. Participants were instructed, "Please read each word aloud as quickly and accurately as possible." These formant values were later used to construct individual vowel spaces for calculating the magnitude and direction of formant changes.

We then established a baseline color-naming latency. Red, green, and blue circles were shown 30 times each. Participants were instructed, "Please name the COLOR aloud as quickly and accurately as possible." These data were not analyzed in the current study.

The main experimental task was performed last. It comprised congruent, vowel-incongruent, and color-incongruent trials, as described above, presented in

random order. There were 40 trials of each of three congruent stimuli, 40 trials of each of six vowel-incongruent stimuli, and 20 trials of each of six color-incongruent stimuli, for a total of 480 trials. Participants were instructed, "Please name the COLOR aloud as quickly and accurately as possible." Short breaks were given after every 48 trials.

As part of a larger Stroop experiment with the same structure, participants in the replication sample encountered trials identical to a subset of those presented to the original sample, namely: the words *red*, *rid*, *rad*, *green*, and *blue* written in red, all requiring the verbal response "red." These results are presented alongside those from the original sample for ease of comparison. Full details of the replication experiment are provided in the Appendix under Supplementary Methods.

## Analysis

### Errors

Error trials were defined as those in which the participant's first utterance was not the full, undistorted name of the stimulus color. Common error types included interference (reading the word instead of naming the color), distortion (laughing or coughing during production), vocal hesitation ("uh blue"), restarting ("r red"), hybridization ("gred"), and self-correction ("bl green"). Error trials were identified by trained research assistants and removed from each participant's data prior to reaction-time and acoustic analysis. To measure the effect of experimental condition on error rates, participants' percentages of error trials were analyzed in a repeated-measures analysis of variance (ANOVA) via the *rstatix* package (Kassambara, 2021) in R (R Core Team, 2021). The measure of effect size was partial eta squared ( $\eta_p^2$ ). Where appropriate, post-hoc paired *t*-tests were conducted with the Bonferroni correction for multiple comparisons and Cohen's *d* as a measure of effect size.

### Reaction time

The purpose of the reaction-time analysis was twofold: to ensure that the color-incongruent condition replicated the classic Stroop effect of longer latencies as compared to the congruent condition, and to determine whether the novel vowel-

incongruent condition also produced chronometric evidence of interference.

Reaction time was calculated as the latency of voicing onset with respect to visual stimulus onset on each trial. To accommodate the skewed distribution of raw reaction times without transformation, we modelled reaction time as an inverse Gaussian distribution and applied a generalized linear mixed-effect model (Bates et al., 2015; Lo & Andrews, 2015) with a fixed factor of condition, random intercepts for participants, and random by-participant slopes for condition. Significance was evaluated with a type-III Wald chi-square ( $\chi^2$ ) test (Fox & Weisberg, 2019), and post-hoc tests were conducted using Tukey contrasts via the *multcomp* package (Hothorn et al., 2008) in R.

### Formants

The first and second formant frequencies (F1 and F2) were tracked using *wave\_viewer* (Niziolek & Houde, 2015), a MATLAB interface to Praat software (Boersma & Weenink, 2021). Pre-emphasis values and linear predictive coding order were defined for each participant and color name; errors in formant tracking were resolved by manually adjusting these parameters on a trial-by-trial basis.

Two procedures were used to extract formant values. The first procedure captured all values between the onset and the offset of voicing in each trial, such that data across participants and trials were aligned to voicing onset at time = 0 ("voicing-aligned"). Voicing boundaries were marked by trained research assistants, blind to condition, using landmarks in the waveform and spectrogram (i.e., visible periodicity) and in the amplitude trace. The second procedure used timestamps for the onset and offset of each phone, generated by the Montreal Forced Aligner (McAuliffe et al., 2017), to align the formant data to vowel onset at time = 0 ("vowel-aligned"). Given that the onset of voicing occurs, with variable timing, within the /ɪ/ segments of "red" and "green" and the /ɪ/ segment of "blue", this second procedure assesses which effects specifically occur in the vowel. Voicing-aligned results are reported in the main text. Vowel-aligned results are summarized in the main text, with figures provided in the Appendix. Except where stated otherwise, results refer to voicing-aligned data.

Extracted formant trajectories for each trial were then zero-phase filtered with an eight-point Hamming window and converted to the logarithmic mel frequency scale, a perceptually-based alternative to the Hertz scale based on human sensitivity to pitch changes (O'Shaughnessy, 1987). Except where noted otherwise, all data processing and analysis were performed in MATLAB 2020b (Mathworks, Natick, MA).

To identify the acoustic effects of incongruent information during color naming, the average F1 and F2 frequencies were first calculated at each time point across all congruent trials, producing a baseline trajectory for each participant and color word. This congruent baseline average was then subtracted from each corresponding vowel-incongruent trial and color-incongruent trial (i.e., the formants of "red" were only compared to those of other productions of "red" under different conditions). These values formed the **F1 difference** and **F2 difference** metrics, representing how a color word is pronounced differently in the presence of incongruent information.

To translate F1 and F2 differences into movement in two-dimensional vowel space, we first defined vectors between the stimulus words as produced by each participant in the neutral context. To obtain an estimate of these vowels' steady-state formants, we used the median F1 and F2 values taken from the middle 50 percent of each voicing-aligned formant track. Then, for each incongruent trial and at each time point, we collapsed the two formant differences into a single deviation by calculating the Euclidean distance from baseline to production in F1-F2 space. Finally, to quantify the degree to which this production constituted an attraction vs. a repulsion by the distractor, we computed the scalar projection of this deviation vector onto the neutral target-to-distractor vector (Niziolek & Guenther, 2013). The component of this formant deviation that lies along the neutral vector has a positive value and indicates **attraction**, corresponding to an acoustic bias towards the distractor. When this metric takes on a negative value, it indicates **repulsion** — an acoustic bias away from the distractor — such that a component of the deviation lies along the *inverse* of the neutral vector. When this metric takes on a value of 0, it indicates no formant bias or a bias that is orthogonal to the neutral vector.

Significance of the F1 difference, F2 difference, and attraction/repulsion metrics was evaluated across

time points (3-ms resolution) with control of the false discovery rate (corrected  $p < 0.05$ ; Benjamini & Hochberg, 1995). Specifically, we took the following three-step approach: (1) We tested whether the incongruence manipulations had *any* significant effect on formant values. As F1 difference and F2 difference already represented distances from baseline, we performed trialwise two-tailed, one-sample  $t$ -tests on each time point of the difference trajectories associated with each distractor. (2) We then tested whether F1 and F2 differences were *distractor-specific*. If different distractors are associated with different formant changes, it suggests a linguistic effect of the distractor rather than a nonspecific effect of incongruent information on speech production. Here we conducted trialwise two-tailed, two-sample  $t$ -tests on the distractors' difference trajectories within each color and condition: *rid* vs. *rad* during "red", *grin* vs. *grain* during "green", *bleed* vs. *blow* during "blue" (vowel-incongruent condition); *green* vs. *blue* during "red", *red* vs. *blue* during "green", *red* vs. *green* during "blue" (color-incongruent condition). (3) Finally, we tested for reliable attraction/repulsion using trialwise two-tailed, one-sample  $t$ -tests. Biases towards or away from the distractor were defined as statistically significant positive or negative values, respectively, of the attraction/repulsion distance.

### Duration

In order to choose an appropriate time window for formant analysis, we first determined the average syllable duration, measured from voicing onset to voicing offset, across all correct trials for each spoken color name: "red": mean = 229 ms, SD = 63 ms; "green": mean = 264 ms, SD = 76 ms; "blue": mean = 279, SD = 84 ms. We chose to analyze the first 200 ms of these utterances so that the majority of trials ("red": 65%; "green": 79%; "blue": 80%) would contribute data to every time point in the test window. For vowel-aligned data, where vowel onset within the voiced syllable is fixed at time = 0, the test window was shifted to -50–150 ms.

Because we planned to analyze the effect of condition on formant trajectories over time, we then assessed whether duration varied across the congruent, vowel-incongruent, and color-incongruent conditions. A significant effect of condition was identified within each color name ("red":  $F(2, 30.72) = 6.22$ ,  $p = 0.005$ ; "green":  $F(2, 97.37) = 6.72$ ,  $p =$

0.002; “blue”:  $F(2, 26.16) = 5.70, p = 0.009$ ). (Each was analyzed in a linear mixed-effects model with a fixed factor of condition, random intercepts for participants, and random by-participant slopes for condition.) Although statistically significant, these differences in mean duration by condition were on the order of 6–10 ms, and so we elected to analyze the raw trajectories in the main text. However, in order to increase the reliability of effects across the time window, we performed ancillary analyses using duration-normalized formant trajectories. Each F1 and F2 trajectory was normalized by sampling it at every two percent of its duration, thereby fixing each duration at 50 samples. These results are summarized in the main text, with figures provided in the Appendix.

### Acoustic variability

An alternative outcome for this study was that, rather than causing a systematic bias in pronunciation, cognitive interference would have an unsystematic, destabilizing effect on speech motor performance. To this end, we investigated whether metrics of acoustic variability varied by experimental condition.

For **formant variability**, we first averaged the F1 and F2 values from the middle 50 percent of each voicing-aligned utterance. We then identified the median F1 and F2 value within each participant, condition, and color name, and calculated each trial’s distance in F1-F2 space from that median — the centre of its distribution. The trialwise distances were submitted to a linear mixed-effects model (Bates et al., 2015) with a fixed factor of condition, random intercepts for participants, and random by-participant slopes for condition. This analysis tests whether pronunciation is more or less variable under greater cognitive load.

To parallel the analysis of systematic vs. unsystematic formant changes by condition, we also investigated systematic vs. unsystematic changes in vocal pitch. Independent of its potential effect on pronunciation via articulatory planning, the cognitive effort needed to resolve interference may cause an increase in autonomic arousal that impacts laryngeal function and therefore phonation. The fundamental frequency ( $f_0$ ) of the speech signal has been found to increase with increasingly difficult Stroop tasks (Van Segbroeck et al., 2014; but cf. MacPherson et al., 2017). For **mean  $f_0$** , we averaged  $f_0$ , as calculated by wave\_viewer,

from the middle 50 percent of each voicing-aligned utterance and submitted the trialwise  $f_0$  values to a linear mixed-effects model with a fixed factor of condition, random intercepts for participants, and random by-participant slopes for condition. This analysis tests whether the three Stroop conditions, reflecting various levels of cognitive-linguistic challenge, have systematic effects on vocal pitch.

Finally, participants may produce a greater range of  $f_0$  values under stressful circumstances (Wittels et al., 2002; but cf. Scherer et al., 2002). For  **$f_0$  variability**, we first averaged the  $f_0$  values from the middle 50 percent of each voicing-aligned utterance. We then identified the median  $f_0$  value within each participant, condition, and color name, and calculated each trial’s difference from that median. The trialwise differences were submitted to a linear mixed-effects model with a fixed factor of condition, random intercepts for participants, and random by-participant slopes for condition. This analysis tests whether vocal pitch is more or less variable under greater cognitive load.

For all three outcome variables, significance at the 0.05 level was evaluated in a type-III ANOVA using the *lmerTest* package (Kuznetsova et al., 2017), and  $p$ -values were approximated with Satterthwaite’s method for calculating degrees of freedom.

## Results

### The color-word interference effect

Error and reaction-time analyses confirmed that color-word incongruence produced interference during color naming. The percentage of error-free trials significantly differed by condition,  $F(1.07, 17.13) = 25.12, p < 0.001, \eta_p^2 = 0.61$ . This percentage was significantly lower in the color-incongruent condition (mean = 91.52%) as compared to the congruent condition (mean = 99.26%;  $t(16) = 4.96, p < 0.001, d = 1.20$ ) and to the vowel-incongruent condition (mean = 99.12%;  $t(16) = 5.20, p < 0.001, d = 1.26$ ), indicating that participants made more frequent speech errors in the color-incongruent condition. Error trials were discarded from further analysis, leaving an average of 119 (out of a possible 120) trials per participant ( $SD = 2$ , range = 113–120) in the congruent condition; 238 (out of a possible 240) trials ( $SD = 3$ , range = 228–240) in the vowel-incongruent condition; and 110

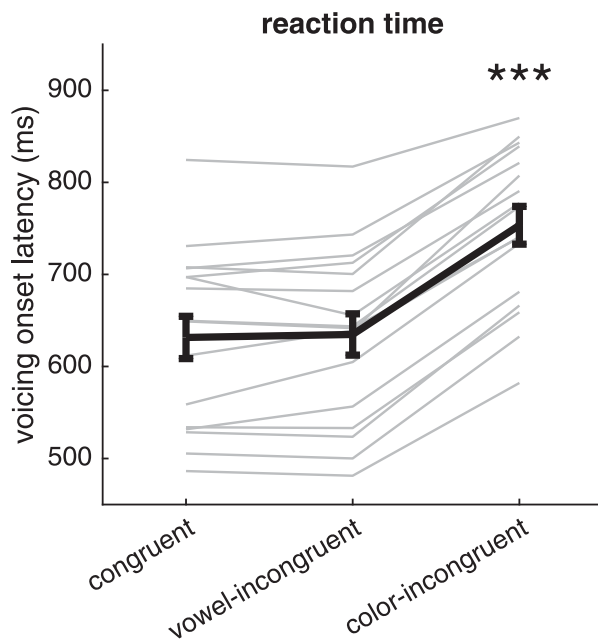


**Figure 2.** Reaction time. Thin lines depict data from individual participants and the thick line depicts the mean and standard error.

(out of a possible 120) trials ( $SD = 9$ , range = 87–120) in the color-incongruent condition.

As shown in Figure 2, reaction times also significantly differed by condition, Wald  $\chi^2(2) = 490.63$ ,  $p < 0.001$ . Verbal responses were 121 ms slower in the color-incongruent condition relative to the congruent condition ( $p < 0.001$ ). Reaction times in the vowel-incongruent and congruent conditions did not differ significantly ( $p = 0.77$ ).

### Effects of vowel incongruence on formants

On vowel-incongruent trials, participants were tasked with naming the text color (“red”, “green”, or “blue”) when the written distractor mostly overlapped in pronunciation but differed in vowel nucleus: *rid* and *rad*; *grin* and *grain*; *bleed* and *blow*. We took a three-step approach to evaluate effects of the distractor on the pronunciation of the target utterance, culminating in an analysis of whether the target is attracted by or repelled from the acoustic space of the distractor. Because of their near-linear arrangement in formant space (Figure 1(b)), we expected that the stimulus set of *rid*, *red*, and *rad* would provide the clearest evidence for attraction vs. repulsion.

First, we measured the acoustic effect of vowel-incongruent information. We hypothesized that the

F1 and/or F2 of the target word would differ between congruent and vowel-incongruent conditions; e.g., the target “red” would be pronounced differently in the presence of the written word *red* vs. *rid*. Figure 3(a) shows the average difference over the time course of the syllable for each distractor word relative to its congruent-word baseline; significant differences are indicated by horizontal lines in corresponding colors at the bottom of each panel. Overall, we identified significant differences from baseline for five out of six distractors. In both the original and replication samples, *rid* significantly altered the F1 of “red” and *rad* significantly altered both the F1 and the F2 of “red”. *Grain* significantly altered the F2 of “green”. *Bleed* significantly altered the F1 and *blow* significantly altered both the F1 and the F2 of “blue”. These results indicate that the presence of incongruent vowel information influences the acoustic properties of speech.

We next investigated whether these effects on F1 and F2 were distractor-specific, indicating a linguistic influence of the distractor, rather than a general effect of incongruence. In particular, because the target “red” lies between *rid* and *rad* in formant space, varying largely along the F1 dimension, we hypothesized that its F1 would be biased in opposite directions for these two distractors. As indicated by the horizontal black lines at the bottom of each panel in Figure 3(a), the F1 of “red” was significantly different in the presence of *rid* vs. *rad*, and this was true in both participant samples. The F2 of “blue” was significantly different in the presence of *bleed* vs. *blow*, which are distinguished primarily by F2.

Having found evidence that incongruent information can bias speech acoustics in a word-specific manner, we next addressed the central question of this paper: are the acoustic properties of the target utterance attracted by or repelled from the acoustic space of a vowel-incongruent distractor? Figure 3(b) shows a standardized F1-F2 vowel space centred on the congruent-condition target color word, with radii representing the average magnitude and direction of productions in the presence of each distractor. As shown by the location of the distractors in this space (thin arrows pointing outward), *rid* and *rad* appeared to repel the formants of “red”, particularly in the F1 dimension. In other words, the presence of *rid*, which has a lower F1 than “red”, is associated with the production of a higher F1 — and vice versa

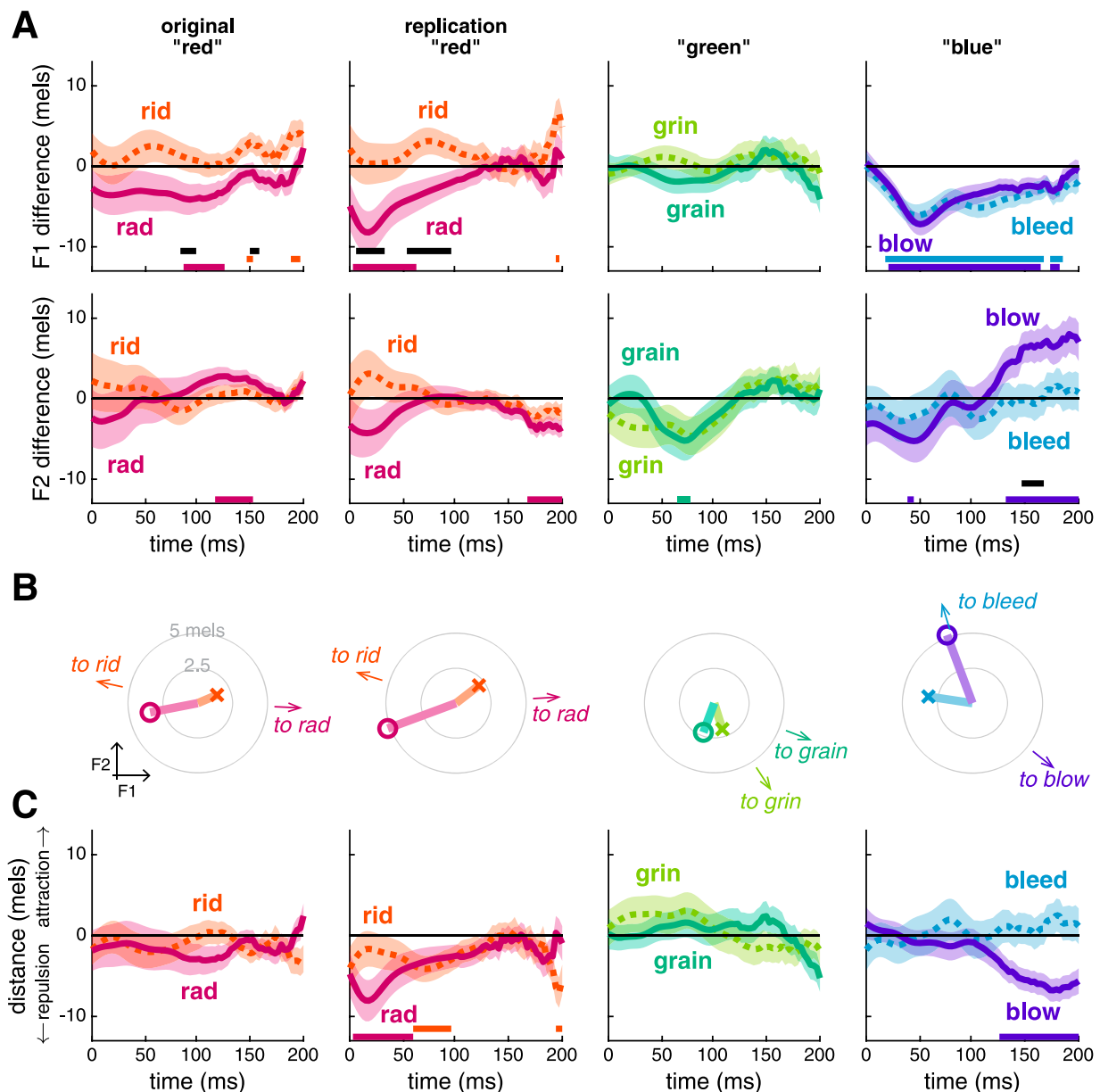


**Figure 3.** Differences in formant trajectories as a function of vowel-incongruent distractor words. Columns, from left to right, show data from spoken responses “red” (original sample), “red” (replication sample), “green”, and “blue”. (A) Mean trajectories over time relative to a congruent-condition baseline are shown for formants F1 (top) and F2 (bottom). Shaded regions depict the standard error of the trialwise mean. (B) Markers (x and o) represent the average formant change, in F1-F2 space, associated with each distractor, relative to the congruent-word origin (0,0). Arrows point to the average location, in the same F1-F2 space, of the distractor words pronounced in a neutral context. Marker data represent averages over time, roughly corresponding to the 100-ms time window of greatest effect: “red”: 0–100 ms; “green”: 50–150 ms; “blue”: 100–200 ms. (C) Time courses of formant attraction (positive values) vs. repulsion (negative values) by each distractor. *Statistical significance:* Horizontal lines at the bottom of panels in (A) and (C) indicate time points at which the trace of the corresponding color differs significantly from 0; horizontal black lines indicate time points at which the two trajectories differ significantly from one another.

for the presence of *rad*: its relatively higher F1 is associated with the production of a lower F1. These patterns were observed in both the original and the replication samples, indicating that formants tend to become *less* like those of a salient written distractor. Effects in vowel-incongruent “green” trials appeared

smaller and less consistent. “Blue” trials evidenced formant changes orthogonal to or away from the distractor. In Figure 3(c), we quantified attraction/repulsion for each distractor word separately by computing, at each time point in the syllable, the component of the formant change in F1-F2 space in

the direction of each distractor. In other words, we computed the scalar projection onto the vector between the target and the distractor (see Figure 1(b) for the group-average vector), such that positive distances indicate attraction and negative distances indicate repulsion. The magnitude of this distance reached significance for several distractors (original sample: *blow*; replication sample: *rid* and *rad*), all of which showed a significant repellent effect on target formants. No distractors significantly attracted target formants.

### Duration-normalized analyses

Because of small but significant differences in the duration of utterances by condition, we repeated the above analyses with duration-normalized formant trajectories in order to improve reliability across the entire test window. The overall pattern of results was largely maintained (Figure S1). In particular, both participant samples still showed distractor-specific effects on the F1 of “red”, which corresponded to significant repulsion of “red” by *rad* in the replication sample. There were no significant effects for “green”. For “blue”, results were as described above, with additional effects of *bleed*, which had a significant repellent effect on “blue” and which differed from *blow* in F1 as well as F2.

### Vowel-aligned analyses

Finally, to assess which formant effects specifically occurred during the vowel, we repeated our analyses on trajectories aligned to vowel onset at time = 0 (Figure S2). This procedure demonstrated that significant and distractor-specific effects on the F1 of “red” occurred during both the /ɪ/ and /ɛ/ segments. Additionally, it revealed that *rid* and *rad* had significant and specific effects on F2 during the vowel segment, which were not apparent in the voicing-aligned data. Vowel-alignment confirmed significant formant repulsion by *rid* and *rad* both prior to and during the vowel (original sample) and by *rad* prior to the vowel (replication sample). For “green”, alignment revealed that *grin* altered F2 prior to the vowel and *grain* altered F2 during the vowel. This corresponded to dynamic effects of *grin*, attracting “green” prior to the vowel and repelling it near the end of the vowel. For “blue”, significant formant changes were revealed to occur both prior to and

during the vowel, with the distractor-specific effects in F2 occurring during the vowel. The sustained repulsion of “blue” by *blow* spanned the /ɪ/ and /u/ segments. In sum, the vowel-aligned analyses were consistent in showing largely repellent effects across multiple distractors.

### Effects of color incongruence on formants

On color-incongruent trials, participants were tasked with naming the text color (“red”, “green”, or “blue”) of the written word when that word was, itself, a valid color-name response on other experimental trials. As evidenced by higher error rates and slower reaction times, the presence of a color-incongruent distractor created substantial interference during color naming. We performed the same formant analyses on these data to determine whether this classic Stroop manipulation affected the pronunciation of the target words. Figure 4(a) shows the average F1 and F2 differences over the time course of the syllable for each distractor word relative to the same syllable produced in the congruent condition. All six color-incongruent distractors yielded statistically reliable differences in the formant trajectories of the target words. These differences were evident in both F1 and F2 and were notably of a larger magnitude than the vowel-incongruent effects. Distractor-specific effects (horizontal black lines in Figure 4(a)), in which the two distractor-influenced trajectories were significantly different from one another, were evident in “red” (replication sample) and “green”; in both targets, both F1 and F2 were significantly different between the two distractors.

The classic Stroop distractor words, in addition to generating interference, differed greatly from the target words in their phonological form. We asked whether, under these conditions of little phonological overlap, the observed formant changes corresponded to an increase or a decrease in similarity of target and distractor vowels. In Figure 4(b) (as in Figure 3(b)), patterns of attraction/repulsion are illustrated in a standardized vowel space that is centred on the congruent production, and Figure 4(c) (like Figure 3(c)) shows only the component of formant change that aligns with (attraction) or opposes (repulsion) the target-distractor vector. Again, these effects tended to be larger than

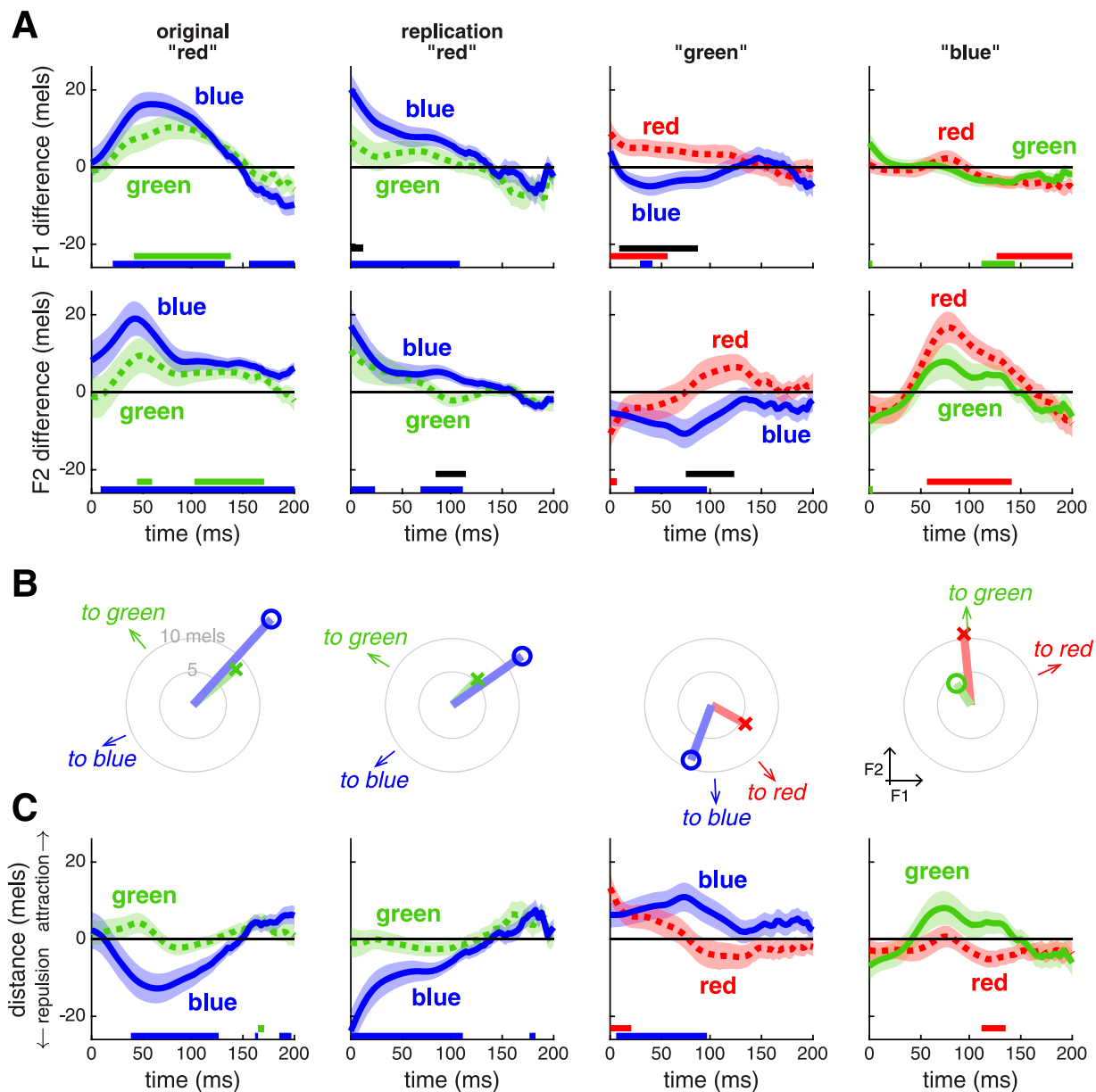


**Figure 4.** Differences in formant trajectories as a function of color-incongruent distractor words. Columns, from left to right, show data from spoken responses "red" (original sample), "red" (replication sample), "green", and "blue". (A) Mean trajectories over time relative to a congruent-condition baseline are shown for formants F1 (top) and F2 (bottom). Shaded regions depict the standard error of the trialwise mean. (B) Markers (x and o) represent the average formant change, in F1-F2 space, associated with each distractor, relative to the congruent-word origin (0,0). Arrows point to the average location, in the same F1-F2 space, of the distractor words pronounced in a neutral context. Marker data represent averages over time, roughly corresponding to the 100-ms time window of greatest effect: "red": 0–100 ms; "green": 0–100 ms; "blue": 50–150 ms. (C) Time courses of formant attraction (positive values) vs. repulsion (negative values) by each distractor. *Statistical significance as in Figure 3.*

those in the vowel-incongruent condition (note scale). While the vowel-incongruent condition produced only significant repellent effects, some targets in the color-incongruent condition underwent significant attraction; others, significant repulsion. Effects on "red" were similar in magnitude and direction in the two samples, with sizable repulsion by *blue* and orthogonal movement by *green*; effects

on "green" were largely attractive; and "blue" demonstrated orthogonal or minor (but significant) repellent effects.

#### *Duration-normalized analyses*

Normalizing the duration of all trials' trajectories had minor effects upon these results (Figure S3). Significant repellent effects were observed for five of the

eight stimuli; “green” was the only case in which significant formant attraction was observed.

### Vowel-aligned analyses

By aligning trajectories to vowel onset at time = 0, we observed that effects were present both prior to and during the vowel in each utterance (Figure S4). For “red”, the replication sample tended to show earlier effects than the original sample. Several effects on “green” were also identified prior to the vowel, including the attraction by both distractors. For “blue”, the formant differences and repulsion were largely confined to the vowel. In sum, the vowel-aligned analyses were consistent in showing significant repellent effects for “red” and “blue”, and attractive effects for “green”.

### Summary of formant changes induced by incongruent distractors

Results for all vowel-incongruent distractors and both samples are summarized in Figure 5(a), where the component of formant change in the direction of each distractor (i.e., the scalar projection into the target-distractor vector, as in Figure 3(c)) is plotted on the y-axis; positive values indicate attraction and negative values indicate repulsion. While effects differed across stimuli, formant changes were largely consistent with a repulsion from the acoustic space of the distractor. Of note, repulsion of “red” by *rid* (orange x’s) and by *rad* (magenta o’s) was consistent

across the two samples. The magnitude of repulsion reached significance at the item level for three distractors (original sample: *blow*; replication sample: *rid* and *rad*). None of the distractors significantly attracted formants.

Results for all color-incongruent distractors and both samples (as in Figure 4(c)) are summarized in Figure 5 (b). Colors indicate distractor identity and labels indicate the target word. Here, too, effects on “red” were similar in the two samples. There was significant bidirectional repulsion between *red* and *blue* – “red” was repelled by *blue* (in both samples), and vice versa – and significant attraction of “green” by both distractors.

### Acoustic variability

We evaluated the hypothesis that the presence of incongruent information might disrupt speech planning and performance in a general sense, resulting in more variable acoustic realizations. However, neither formant variability ( $F(2, 25.93) = 1.59, p = 0.22$ ) nor  $f_0$  variability ( $F(2, 26.26) = 1.33, p = 0.28$ ) differed across the three conditions. Additionally, condition had no systematic effect on the mean  $f_0$  of the color-word responses,  $F(2, 31.67) = 1.16, p = 0.33$ .

### Discussion

The primary aim of this study was to determine if and how a competing vowel representation would affect

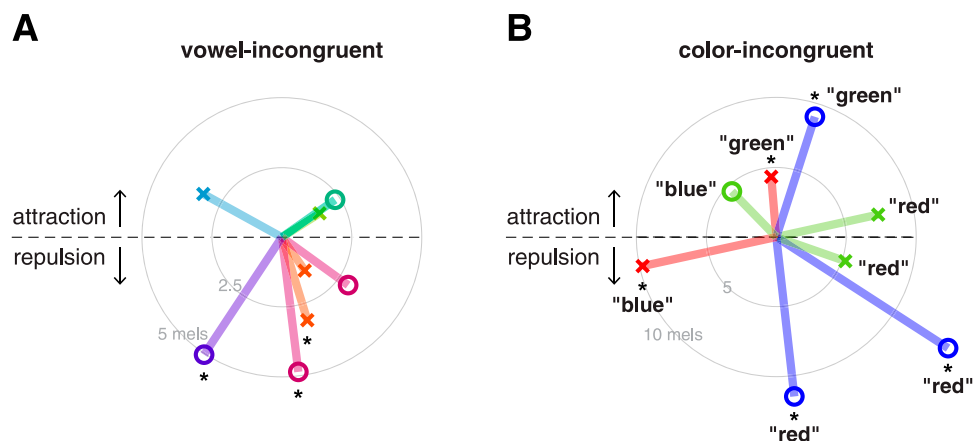


**Figure 5.** Summary of attraction/repulsion by distractors. To summarize the formant movement of targets with respect to distractors, the component of formant change in the direction of each distractor (i.e., the scalar projection onto the target-distractor vector) is plotted on the y-axis, where positive values constitute attraction and negative values constitute repulsion (as in Figures 3(c) and 4(c)). The perpendicular component of formant change (i.e., the scalar projection onto the orthogonal vector) is plotted on the x-axis. “Red” data from both samples are shown. (A) Vowel-incongruent summary data. Colors indicate distractor identity. (B) Color-incongruent summary data. Colors indicate distractor identity and labels indicate the target word. *Statistical significance* is indicated by an asterisk.

the acoustic properties of speech. We developed a vowel-modified Stroop task in which participants were tasked with, for example, naming the color “blue” in the presence of the written word *blow*. We found that the presence of such a distractor in many cases significantly altered formant trajectories, and that these changes primarily repelled the target from the distractor in acoustic vowel space. These results suggest a phenomenon in which cognitive inhibition is reflected in speech motor execution during the resolution of interference.

### **Cognitive inhibition enhances acoustic contrast**

The maintenance of and perceptual preference for acoustic contrast among spoken words has been documented at multiple levels of linguistic analysis. Broadly, phonological systems seem to obey the “principle of maximal contrast” (Lindblom, 1986, p. 15): nearly all languages, including those with only three vowels, have /i/, /a/, and /u/, which occupy the extremes of vowel space. Perceptually distinct contrasts such as these support the efficient transfer of information from speaker to listener (Flemming, 2004). As phoneme inventories change over time, phoneme pairs that distinguish more vs. fewer minimal-pair words are less likely to undergo merger (Wedel et al., 2013), highlighting the pressure on phonological systems to preserve meaningful distinctions. Speakers may accomplish this by selectively shifting certain pronunciations, hyperarticulating words with more vs. fewer phonological neighbours (Wright, 2004) and/or those with vs. without a minimal-pair neighbour (Wedel et al., 2018) in order to enhance distinctiveness. Experimental manipulations of neighbourhood density, competitor salience, and listener feedback are all effective at driving speakers to produce less confusable pronunciations (Baese-Berk & Goldrick, 2009; Buz et al., 2016; Fox et al., 2015; Stern & Shaw, 2023).

The current study differs somewhat from previous reports of adaptive phonetic contrast in that we leveraged the Stroop effect in order to automatically activate a competing vowel representation. Therefore, the “attraction” hypothesis — that these coactivated representations would yield a hybrid pronunciation — was also compelling. For one thing, the contamination of speech by irrelevant environmental language is well documented (Harley, 1984). For

another, active vowel representations can migrate from word to word (e.g., the /a/ of *might* is anticipated in the speech error “one spice — spouse might say ...” (Cutler, 1988)). Moreover, *gradient* errors rather than full substitutions are readily elicited in high-conflict situations such as speeded tongue-twister recitation (Frisch & Wright, 2002; Goldrick & Blumstein, 2006; Goldrick et al., 2016; Goldstein et al., 2007; Mowrey & MacKay, 1990), as well as in spontaneous speech (Alderete et al., 2021). Trace phonetic effects have been attributed to phonological coactivations cascading into articulatory processes, such that errorful speech sounds reflect characteristics of the intended target (Goldrick & Blumstein, 2006). Applied to our study, the “attraction” hypothesis would predict a complementary effect, that *correct* speech sounds would reflect characteristics of a salient competitor. However, the processes elicited in the current study were more consistent with an inhibitory mechanism that *prevents* such an intrusion and ultimately increases the acoustic distance between target and distractor.

Interference from vowel-incongruent distractors such as *rid* and *rad* may be expected to be low: they are not colors or even semantically related to colors, and accordingly lack response relevance (Lupker, 1979). However, several studies have reached the conclusion, on the basis of chronometric data, that vowel production necessarily involves the inhibition of neighbouring vowels. For example, Yaniv and colleagues (1990) found that response latencies were longer for spoken pairings of similar vowels (e.g., “peat-pit”, “putt-pot”) as compared to dissimilar vowels (e.g., “pit-pot”, “pit-putt”), even when these utterances were anticipated, suggesting that more similar vowels generated more interference. Wheel-don (2003) documented a case of inhibitory priming whereby picture naming was slowed by the production, on the previous trial, of a phonological neighbour (e.g., Dutch words “blad” and “bloed”, “bus” and “bos”). (Unfortunately, neither study measured whether juxtaposed vowels became more similar or more different to one another.) It is notable, then, that we did not find longer reaction times in our vowel-incongruent condition than in the congruent condition. A number of underlying temporal effects may have been at play, perhaps in combination. In terms of temporal facilitation, response retrieval may have been speeded by the

phonological overlap of the distractor with the target (Marmurek et al., 2006; Jescheniak & Schriefers, 2001); participants may also have been able to begin articulating with only part of the utterance (the initial overlap) planned (Kello et al., 2000). On the other hand, in terms of temporal cost, we could speculate based on Wheeldon's (2003) results that time is required to suppress the phonological encoding of the distractor vowel representation. Here it is important to underscore that the paradigms of Yaniv and colleagues (1990) and Wheeldon (2003) actually elicited the sequential production of competitors, whereas in our task, the vowel-incongruent distractors were present on screen but not to be uttered, and indeed were not potential responses at all (as color-word distractors are in the classic color-incongruent condition). In the words of O'Seaghdha and colleagues (1992, p. 403), "production has the advantage of unambiguously requiring full phonological preparation of primes prior to their articulation". Therefore, our null reaction-time effect may reflect a weaker form of phonological activation driven by reading vs. saying the competitor word, one that does not incur the full temporal cost of disengaging from a phonological-into-motor cascade. However, we caution that our study was not designed to tease apart these temporal effects. Rather, our analysis of formant trajectories demonstrates a competitor's influence at the level of a speech motor programme.

One biologically plausible mechanism driving formant repulsion is lateral inhibition. Commonly described as a feature of sensory systems that increases sensitivity, lateral inhibition is a process by which the firing of one neuron is suppressed by the firing of a neighbouring neuron. In the motor system, lateral inhibition has been invoked to explain programme selection (Sohn & Hallett, 2004), and the role of circuit inhibition in controlling voluntary and precise motor output is well recognized (Duque et al., 2017; Mink, 1996). As speaking is perhaps the most complex human motor behaviour, lateral inhibition underlies many theoretical and computational accounts of speech motor planning and selection (Bohland et al., 2010; Meyer & Gordon, 1985; Schade & Berg, 1992; Tilsen, 2013; Yaniv et al., 1990). Earlier conceptualizations tended to focus on linguistic features as nodes in a network: preparing to produce a phoneme activates its feature nodes

(e.g., place of articulation, voicing), which in turn inhibits phonemes sharing those features. This helps to explain experimental results in which preparing one utterance slows the production of other, similar utterances (Meyer & Gordon, 1985). In vowel production, where features are relative (e.g., height, backness, roundedness), vowel similarity in terms of phonetic distance has been identified as a critical variable governing interference between production plans (Yaniv et al., 1990).

A continuous feature space representing one's speech-sound inventory provides a plausible interface between neuromotor programmes and cognitive operations such as inhibition and attention. Tilsen (2013) and Stern and Shaw (2023) have proposed that dissimilatory phonetic effects, such as the formant repulsion observed in the present study, could be explained by the inhibition of motor plans that involve overlapping populations of neurons (Erlhagen & Schöner, 2002; Georgopoulos et al., 1986; Guenther, 1995; Houghton & Tipper, 1999). When targets and competitors are neighbours in this feature space, top-down suppression of the competitor representation serves to repel the target representation by shifting the population-level coding of the target. Analogous results, in which continuous movement trajectories veer away from competitors' locations, have been obtained in oculomotor control (Sheliga et al., 1994; Van der Stigchel & Theeuwes, 2008) and arm-reaching paradigms (Tipper et al., 1997). In the case of vowel production, the dimensions of this space likely correspond to articulator position (e.g., tongue height, backness), and thus can be read out as a shift in formant values. Candidate anatomical substrates of vowel-repulsion effects are inferior frontal and ventral premotor regions that are sensitive to, and related to online compensation for, externally-applied formant shifts (Tourville et al., 2008; Niziolek & Guenther, 2013).

We propose that overcoming the prepotency of reading vs. naming in the Stroop framework generates a particularly strong type of competitor suppression that is detectable in the pronunciation of the target utterance — a kind of acoustic inhibition. Dissimilatory phonetic effects have also been obtained in vowels produced subsequent to an incongruent vowel cue (Tilsen, 2009) and in the voice onset time of consonants produced with an on-screen minimal-pair competitor (Baese-Berk & Goldrick, 2009; Stern

& Shaw, 2023), suggesting that acoustic inhibition can operate over time and manifest in various continuous dimensions of speech behaviour. In a step towards unifying empirical evidence of attraction and repulsion in one framework, Stern and Shaw's (2023) model takes only competitor strength as an input, using its sign to, parsimoniously, account for both gradient errors (excitatory: intended sounds intruding upon errorful sounds) and contrastive hyperarticulation (inhibitory: distancing of correct targets from distractors). Future work is needed to understand whether neurocomputational models of contrastive dissimilation account for the multidimensional nature of vowel production as well as for a single scaled consonantal feature such as voice onset time (e.g., Stern et al., 2022).

### Stimulus-specific effects

It is worth noting that our experimental design was constrained by both the English lexicon and the anatomy of the vocal tract. We prioritized using the basic, automatized color names of “red”, “green”, and “blue”, but this likely limited the magnitude and direction of possible formant changes: “green” and “blue” contain the corner vowels /i/ and /u/, which makes repulsion hard to achieve without resorting to extreme articulatory positions. Nevertheless, *bleed* and *blow* had sustained effects on the F1 of “blue”, and *blow* evinced a strong repellent effect. We note that the final consonant of distractor *bleed* somewhat complicates interpretations because that stimulus activated a conflicting syllabic frame in addition to a conflicting vowel (Berent & Marom, 2005). *Grin* and *grain* were, in terms of formant proximity, the best distractors for “green”, but they were not easily distinguished from one another in vowel space and ultimately produced the weakest effects. The generalizability of our findings is certainly limited by the heterogeneity of results across the three color words; follow-up experiments using paradigms with less restrictive stimulus sets (e.g., picture-word interference) would be valuable in demonstrating a generalization of this effect.

“Red”, however, occupies a centralized position in vowel space, allowing for attraction, repulsion, or orthogonal movement within the range of typical articulation. Given their near-linear arrangement in the F1 dimension (which has an inverse relationship with tongue height), the stimulus set of *rid*, *red*, and

*rad* was the strongest test in our data of the attraction vs. repulsion hypotheses. Results were consistent with repulsion: in both participant samples, F1 trajectories were biased in opposite directions, reducing the similarity of target and distractor. Aligning formant trajectories to vowel onset (Figure S2) revealed that the dissimilatory effect on the F1 of “red” was largely evident during the /ɪ/ segment of the utterance, while changes during the vowel were largely observed in F2, not F1. This is perhaps surprising, given that the onset consonant is shared among target and distractors. However, it may also be explained by the coarticulatory influence of the subsequent vowel, which distinguishes the formants of /ɪ/ before vowel onset. We explored this by comparing the raw formant trajectories for the words “rid”, “red”, and “rad” as read aloud by participants in a neutral context (Figure S5). Clear separation of the F1 traces is observed throughout the voiced syllable (Figure S5A) and, crucially, in the /ɪ/ segment prior to vowel onset (time points < 0 in Figure S5B). This indicates that the changes to the consonant segment were consistent with movement away from the realization of that segment in the distractor context, and, more broadly, suggests that cognitive inhibition leaves its trace across planned syllables.

Figure S5 also highlights a difference in dialect between the two participant samples. Although we did not collect detailed demographic information, it is likely that at least a subset of the participants in the replication sample, recruited at the University of Wisconsin–Madison, spoke a regional, Inland-North variant of North American English (Labov et al., 2006). The proximity of “red” and “rad” trajectories is characteristic of the relative /ɛ/-lowering and /æ/-raising in this population (Jacewicz & Fox, 2013). Nevertheless, despite differences in vowel configuration, both the Boston and Madison samples showed significant dissimilatory, repulsive effects on F1. Clear similarities in how the two samples responded to distractors for “red” can be appreciated in Figure 5. Future studies may undertake more controlled and detailed studies relating individual vowel spaces and dialects to the phenomenon of inhibitory vowel repulsion (e.g., Clopper & Tamati, 2014).

### Acoustic effects of color-word interference

Because the majority of relevant previous studies induced competition between minimal pairs, it was

more challenging to formulate specific acoustic hypotheses for the target and distractor words in the classic Stroop color-incongruent condition. As speakers know, simultaneous activation of two semantically related words (frog, toad) can produce phonologically surprising results ("Look, a froad!"). But here, all such blending errors ("gred") were removed from the data, and we were only concerned with what was perceived to be the successful suppression of a salient but irrelevant color word. Remarkably, this process also left subtle phonetic traces.

As expected, color-incongruent distractors were associated with higher error rates and longer reaction times, indicating substantial cognitive interference on these trials. However, we did not find evidence that color incongruence made speech more variable, neither in formant values nor in vocal pitch. Instead, each distractor was associated with significant biases in the pronunciation of the target color name — some constituting formant attraction, others formant repulsion. As with the vowel-incongruent results, formant movement may have been asymmetrically constrained in a word-specific manner: "red", from its central position, was mostly repelled; "green", containing a corner vowel, was attracted inwards; "blue" rose sharply in F2, where it had room to increase. Despite broad similarities in formant movement across the two incongruent conditions, some candidate theories that may explain formant changes in the vowel-incongruent results are less relevant here: as "red", "green", and "blue" are not confusable, there is probably little pressure on speakers to hyperarticulate them. However, it may be the case that because responses in this condition are pronounced more carefully, i.e., as a result of controlled vs. automatic processing, their pronunciations differ from those in the congruent condition in systematic ways (Harnsberger et al., 2008). Perhaps most surprising were the distractor-specific effects on the formants of "red" and "green", suggesting a linguistic influence of even phonologically unrelated distractor stimuli. The sequential cuing framework (Sevald & Dell, 1994) might suggest that the vowels in co-active /ɛd/ and /g.iɪn/ influence one another because they both follow /ɪ/. Both "red" in the presence of *green* and "green" in the presence of *red* showed small but significant attraction, potentially reflecting a gradient form of a

familiar type of speech error — an exchange miscued by shared segments (e.g., *duke and duchess* → *duck and doochess* (Simonini, 1956; see also MacKay, 1970)). On the other hand, aligning formants to vowel onset revealed that the attraction of "green" by *red* occurred prior to the vowel, within the voiced onset of the syllable. Acoustic effects of color-word interference are therefore far from straightforward, and it is still largely unclear how coactivation of non-overlapping competitor phonology may cascade into articulation. In future studies, kinematic data on articulatory gestures during cognitive inhibition would be informative, potentially revealing articulatory dynamics consistent with distractor intrusion or distractor aversion (e.g., Drake & Corley, 2015; McMillan et al., 2009; Pouplier et al., 1999).

### Implications and conclusions

Results from a vowel-modified Stroop task demonstrate that formant trajectories can be modulated by cognitive processes involved in conflict resolution. One motivation for developing such a task was to lay the groundwork for characterizing production impairments in individuals with comorbid communication disorders and executive-control challenges. Persons with aphasia, who have been shown to have deficits in attention and executive control as well as language difficulties, may be particularly susceptible to interference during speech production. Stroop interference effects are larger in aphasic speakers (Wiener et al., 2004; Pompon et al., 2015; Kuzmina & Weekes, 2017), while patients with left frontal lesions *without aphasia* are generally slow to name colors but *not* disproportionately impaired on incongruent Stroop trials (Stuss et al., 2001; but cf. Perret, 1974). Thus, the relationships among language ability, verbal cognitive control, and neuroanatomy are complex.

Speech errors have long been used to probe the nature of production impairments in aphasia. For example, the substitutions, deletions, and transpositions of speech sounds — *phonemic paraphasias* — that are common in aphasic speech could be attributed to deficits in phonological selection. However, Kurowski and Blumstein (2016) identified in these errors phonetic traces of intended sounds — the same sort of gradient errors that typical speakers tend to make in high-conflict contexts. One

interpretation of this finding is that mechanisms of cascading activation in spoken word production are intact in aphasia, but that competitor representations are far more influential. Our approach, focusing on subphonemic competitor effects in *correct* productions, could be complementary, revealing the extent to which persons with aphasia are able to enhance acoustic contrast between intended and interfering words.

In sum, the current study capitalizes on the rich dynamics of speech acoustics to shed light on how the brain manages interference among streams of competing information, and suggests that a phenomenon of *acoustic* inhibition may accompany successful *cognitive* inhibition of a salient competitor during speech production. The finding that inhibited information nonetheless influences how words are pronounced is part of a broader appreciation of the complex, interactive nature of language production.

## Acknowledgements

We thank Swathi Kiran for mentorship and guidance and Ian Quillen for assistance with data collection and processing.

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## Funding

This work was supported by the Dudley Allen Sargent Research Fund and the National Institute on Deafness and Other Communication Disorders [grant number: R00 DC014520].

## ORCID

Sara D. Beach  <http://orcid.org/0000-0003-2144-5837>  
Caroline A. Niziolek  <http://orcid.org/0000-0002-6085-1371>

## References

- Alderete, J., Baese-Berk, M., Leung, K., & Goldrick, M. (2021). Cascading activation in phonological planning and articulation: Evidence from spontaneous speech errors. *Cognition*, 210, 104577. <https://doi.org/10.1016/j.cognition.2020.104577>
- Baese-Berk, M., & Goldrick, M. (2009). Mechanisms of interaction in speech production. *Language and Cognitive Processes*, 24(4), 527–554. <https://doi.org/10.1080/016909060802299378>
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1), 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
- Berent, I., & Marom, M. (2005). Skeletal structure of printed words: Evidence from the Stroop task. *Journal of Experimental Psychology: Human Perception and Performance*, 31(2), 328. <https://doi.org/10.1037/0096-1523.31.2.328>
- Boersma, P., & Weenink, D. (2021). Praat: Doing phonetics by computer. <http://www.praat.org/>.
- Bohland, J. W., Bullock, D., & Guenther, F. H. (2010). Neural representations and mechanisms for the performance of simple speech sequences. *Journal of Cognitive Neuroscience*, 22(7), 1504–1529. <https://doi.org/10.1162/jocn.2009.21306>
- Booth, J. R., Perfetti, C. A., & MacWhinney, B. (1999). Quick, automatic, and general activation of orthographic and phonological representations in young readers. *Developmental Psychology*, 35(1), 3. <https://doi.org/10.1037/0012-1649.35.1.3>
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, 10(4), 433–436. <https://doi.org/10.1163/156856897X00357>
- Buz, E., Tanenhaus, M. K., & Jaeger, T. F. (2016). Dynamically adapted context-specific hyper-articulation: Feedback from interlocutors affects speakers' subsequent pronunciations. *Journal of Memory and Language*, 89, 68–86. <https://doi.org/10.1016/j.jml.2015.12.009>
- Clopper, C. G., & Tamati, T. N. (2014). Effects of local lexical competition and regional dialect on vowel production. *The Journal of the Acoustical Society of America*, 136(1), 1–4. <https://doi.org/10.1121/1.4883478>
- Cutler, A. (1988). The perfect speech error. In *Language, speech and mind: Studies in honor of Victoria A. Fromkin* (pp. 209–223). Croom Helm.
- Dennis, I., & Newstead, S. E. (1981). Is phonological recoding under strategic control? *Memory & Cognition*, 9(5), 472–477. <https://doi.org/10.3758/BF03202341>
- Drake, E., & Corley, M. (2015). Articulatory imaging implicates prediction during spoken language comprehension. *Memory & Cognition*, 43(8), 1136–1147. <https://doi.org/10.3758/s13421-015-0530-6>
- Dromey, C., & Bates, E. (2005). Speech interactions with linguistic, cognitive, and visuomotor tasks. *Journal of Speech, Language, and Hearing Research*, 48(2), 295–305. [https://doi.org/10.1044/1092-4388\(2005/020\)](https://doi.org/10.1044/1092-4388(2005/020))
- Duque, J., Greenhouse, I., Labruna, L., & Ivry, R. B. (2017). Physiological markers of motor inhibition during human behavior. *Trends in Neurosciences*, 40(4), 219–236. <https://doi.org/10.1016/j.tins.2017.02.006>
- Erlhagen, W., & Schöner, G. (2002). Dynamic field theory of movement preparation. *Psychological Review*, 109(3), 545. <https://doi.org/10.1037/0033-295X.109.3.545>

- Flemming, E. (2004). Contrast and perceptual distinctiveness. In B. Hayes, R. Kirchner, & D. Steriade (Eds.), *Phonetically-based phonology* (pp. 232–267). Cambridge University Press.
- Fox, J., & Weisberg, S. (2019). *An {R} companion to applied regression* (3rd ed.). Sage.
- Fox, N. P., Reilly, M., & Blumstein, S. E. (2015). Phonological neighborhood competition affects spoken word production irrespective of sentential context. *Journal of Memory and Language*, 83, 97–117. <https://doi.org/10.1016/j.jml.2015.04.002>
- Frisch, S. A., & Wright, R. (2002). The phonetics of phonological speech errors: An acoustic analysis of slips of the tongue. *Journal of Phonetics*, 30(2), 139–162. <https://doi.org/10.1006/jpho.2002.0176>
- Georgopoulos, A. P., Schwartz, A. B., & Kettner, R. E. (1986). Neuronal population coding of movement direction. *Science*, 233(4771), 1416–1419. <https://doi.org/10.1126/science.3749885>
- Goldrick, M., & Blumstein, S. E. (2006). Cascading activation from phonological planning to articulatory processes: Evidence from tongue twisters. *Language and Cognitive Processes*, 21(6), 649–683. <https://doi.org/10.1080/01690960500181332>
- Goldrick, M., Keshet, J., Gustafson, E., Heller, J., & Needle, J. (2016). Automatic analysis of slips of the tongue: Insights into the cognitive architecture of speech production. *Cognition*, 149, 31–39. <https://doi.org/10.1016/j.cognition.2016.01.002>
- Goldstein, L., Pouplier, M., Chen, L., Saltzman, E., & Byrd, D. (2007). Dynamic action units slip in speech production errors. *Cognition*, 103(3), 386–412. <https://doi.org/10.1016/j.cognition.2006.05.010>
- Guenther, F. H. (1995). Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production. *Psychological Review*, 102(3), 594. <https://doi.org/10.1037/0033-295X.102.3.594>
- Harley, T. A. (1984). A critique of top-down independent levels models of speech production: Evidence from non-plan-internal speech errors. *Cognitive Science*, 8(3), 191–219. [https://doi.org/10.1207/s15516709cog0803\\_1](https://doi.org/10.1207/s15516709cog0803_1)
- Harnishfeger, K. K. (1995). The development of cognitive inhibition: Theories, definitions, and research evidence. In F. N. Dempster, & C. J. Brainerd (Eds.), *Interference and inhibition in cognition* (pp. 175–204). Academic Press.
- Harnsberger, J. D., Wright, R., & Pisoni, D. B. (2008). A new method for eliciting three speaking styles in the laboratory. *Speech Communication*, 50(4), 323–336. <https://doi.org/10.1016/j.specom.2007.11.001>
- Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal*, 50(3), 346–363. <https://doi.org/10.1002/bimj.200810425>
- Houghton, G., & Tipper, S. P. (1999). Attention and the control of action: An investigation of the effects of selection on population coding of hand and eye movement. In D. Heinke, G. W. Humphreys, & A. Olson (Eds.), *Connectionist models in cognitive neuroscience* (pp. 283–298). Springer.
- Jaciewicz, E., & Fox, R. A. (2013). Cross-dialectal differences in dynamic formant patterns in American English vowels. In G. S. Morrison, & P. F. Assmann (Eds.), *Vowel inherent spectral change* (pp. 177–198). Springer.
- Jescheniak, J. D., & Schriefers, H. (2001). Priming effects from phonologically related distractors in picture-word interference. *The Quarterly Journal of Experimental Psychology*, 54(2), 371–382. <https://doi.org/10.1080/713755981>
- Kassambara, A. (2021). rstatix: Pipe-friendly framework for basic statistical tests. R package version 0.7.0. <https://CRAN.R-project.org/package=rstatix>.
- Kello, C. T., Plaut, D. C., & MacWhinney, B. (2000). The task dependence of staged versus cascaded processing: An empirical and computational study of Stroop interference in speech perception. *Journal of Experimental Psychology: General*, 129(3), 340. <https://doi.org/10.1037/0096-3445.129.3.340>
- Kleinow, J., & Smith, A. (2006). Potential interactions among linguistic, autonomic, and motor factors in speech. *Developmental Psychobiology*, 48(4), 275–287. <https://doi.org/10.1002/dev.20141>
- Kurowski, K., & Blumstein, S. E. (2016). Phonetic basis of phonemic paraphasias in aphasia: Evidence for cascading activation. *Cortex*, 75, 193. <https://doi.org/10.1016/j.cortex.2015.12.005>
- Kuzmina, E., & Weekes, B. S. (2017). Role of cognitive control in language deficits in different types of aphasia. *Aphasiology*, 31(7), 765–792. <https://doi.org/10.1080/02687038.2016.1263383>
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13), 1–26. <https://doi.org/10.18637/jss.v082.i13>
- Labov, W., Ash, S., & Boberg, C. (2006). *The atlas of north American English: Phonetics, phonology and sound change*. Walter de Gruyter.
- Lindblom, B. (1986). Phonetic universals in vowel systems. In J. J. Ohala, & J. J. Jaeger (Eds.), *Experimental phonology*. Academic Press.
- Lo, S., & Andrews, S. (2015). To transform or not to transform: Using generalized linear mixed models to analyse reaction time data. *Frontiers in Psychology*, 6, 1171. <https://doi.org/10.3389/fpsyg.2015.01171>
- Lupker, S. J. (1979). The semantic nature of response competition in the picture-word interference task. *Memory & Cognition*, 7(6), 485–495. <https://doi.org/10.3758/BF03198265>
- MacKay, D. G. (1970). Spoonerisms: The structure of errors in the serial order of speech. *Neuropsychologia*, 8(3), 323–350. [https://doi.org/10.1016/0028-3932\(70\)90078-3](https://doi.org/10.1016/0028-3932(70)90078-3)
- MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, 109(2). <https://doi.org/10.1037//0033-2909.109.2.163>
- MacPherson, M. K., Abur, D., & Stepp, C. E. (2017). Acoustic measures of voice and physiologic measures of autonomic arousal during speech as a function of cognitive load. *Journal of Voice*, 31(4), 504-e1–e9. <https://doi.org/10.1016/j.jvoice.2016.10.021>
- Marmurek, H. H., Proctor, C., & Javor, A. (2006). Stroop-like serial position effects in color naming of words and nonwords.

- Experimental Psychology*, 53(2), 105–110. <https://doi.org/10.1027/1618-3169.53.2.105>
- McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., & Sonderegger, M. (2017, August). Montreal Forced Aligner: Trainable text-speech alignment using Kaldi. *Interspeech*, 2017, 498–502.
- McMillan, C. T., Corley, M., & Lickley, R. J. (2009). Articulatory evidence for feedback and competition in speech production. *Language and Cognitive Processes*, 24(1), 44–66. <https://doi.org/10.1080/01690960801998236>
- Meyer, D. E., & Gordon, P. C. (1985). Speech production: Motor programming of phonetic features. *Journal of Memory and Language*, 24(1), 3–26. [https://doi.org/10.1016/0749-596X\(85\)90013-0](https://doi.org/10.1016/0749-596X(85)90013-0)
- Mink, J. W. (1996). The basal ganglia: Focused selection and inhibition of competing motor programs. *Progress in Neurobiology*, 50(4), 381–425. [https://doi.org/10.1016/S0301-0082\(96\)00042-1](https://doi.org/10.1016/S0301-0082(96)00042-1)
- Mousikou, P., Roon, K. D., & Rastle, K. (2015). Masked primes activate feature representations in reading aloud. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 41(3), 636. <https://doi.org/10.1037/xlm0000072>
- Mowrey, R. A., & MacKay, I. R. (1990). Phonological primitives: Electromyographic speech error evidence. *The Journal of the Acoustical Society of America*, 88(3), 1299–1312. <https://doi.org/10.1121/1.399706>
- Niziolek, C. A., & Guenther, F. H. (2013). Vowel category boundaries enhance cortical and behavioral responses to speech feedback alterations. *Journal of Neuroscience*, 33(29), 12090–12098. <https://doi.org/10.1523/JNEUROSCI.1008-13.2013>
- Niziolek, C. A., & Houde, J. (2015). *Wave\_Viewer: First release*. <https://doi.org/10.5281/ZENODO.13839>.
- Oostwoud Wijdenes, L., Ivry, R. B., & Bays, P. M. (2016). Competition between movement plans increases motor variability: Evidence of a shared resource for movement planning. *Journal of Neurophysiology*, 116(3), 1295–1303. <https://doi.org/10.1152/jn.00113.2016>
- O'Seaghdha, P. G., Dell, G. S., Peterson, R. R., & Juliano, C. (1992). Models of form-related priming in comprehension and production. In R. G. Reilly, & N. E. Sharkey (Eds.), *Connectionist approaches to natural language processing*. Routledge.
- O'Shaughnessy, D. (1987). *Speech communication: Human and machine*. Addison-Wesley.
- Perret, E. (1974). The left frontal lobe of man and the suppression of habitual responses in verbal categorical behaviour. *Neuropsychologia*, 12(3), 323–330. [https://doi.org/10.1016/0028-3932\(74\)90047-5](https://doi.org/10.1016/0028-3932(74)90047-5)
- Pompon, R. H., McNeil, M. R., Spencer, K. A., & Kendall, D. L. (2015). Intentional and reactive inhibition during spoken word Stroop task performance in people with aphasia. *Journal of Speech, Language, and Hearing Research*, 58(3), 767–780. [https://doi.org/10.1044/2015\\_JSLHR-L-14-0063](https://doi.org/10.1044/2015_JSLHR-L-14-0063)
- Poupplier, M., Chen, L., Goldstein, L., & Byrd, D. (1999). Kinematic evidence for the existence of gradient speech errors. *The Journal of the Acoustical Society of America*, 106(4), 2242–2242. <https://doi.org/10.1121/1.427638>
- Rapp, B., & Goldrick, M. (2000). Discreteness and interactivity in spoken word production. *Psychological Review*, 107(3), 460. <https://doi.org/10.1037/0033-295X.107.3.460>
- Rayner, K., Sereno, S. C., Lesch, M. F., & Pollatsek, A. (1995). Phonological codes are automatically activated during reading: Evidence from an eye movement priming paradigm. *Psychological Science*, 6(1), 26–32. <https://doi.org/10.1111/j.1467-9280.1995.tb00300.x>
- R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Schade, U., & Berg, T. (1992). The role of inhibition in a spreading-activation model of language production. II. The simulation perspective. *Journal of Psycholinguistic Research*, 21(6), 435–462. <https://doi.org/10.1007/BF01067523>
- Scherer, K. R., Grandjean, D., Johnstone, T., Klasmeyer, G., & Bänziger, T. (2002, September). Acoustic correlates of task load and stress. *Interspeech*, 2002.
- Sevald, C. A., & Dell, G. S. (1994). The sequential cuing effect in speech production. *Cognition*, 53(2), 91–127. [https://doi.org/10.1016/0010-0277\(94\)90067-1](https://doi.org/10.1016/0010-0277(94)90067-1)
- Sheliga, B. M., Riggio, L., & Rizzolatti, G. (1994). Orienting of attention and eye movements. *Experimental Brain Research*, 98(3), 507–522. <https://doi.org/10.1007/BF00233988>
- Simonini, R. C. (1956). Phonemic and analogic lapses in radio and television speech. *American Speech*, 31(4), 252–263. <https://doi.org/10.2307/453412>
- Sohn, Y. H., & Hallett, M. (2004). Surround inhibition in human motor system. *Experimental Brain Research*, 158(4), 397–404.
- Stern, M. C., Chaturvedi, M., & Shaw, J. A. (2022). *A dynamic neural field model of phonetic trace effects in speech errors*. Proceedings of the Annual Meeting of the Cognitive Science Society (Vol. 44, No. 44).
- Stern, M. C., & Shaw, J. A. (2023). Neural inhibition during speech planning contributes to contrastive hyperarticulation. *Journal of Memory and Language*, 132, 104443. <https://doi.org/10.1016/j.jml.2023.104443>
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18(6), 643. <https://doi.org/10.1037/h0054651>
- Stuss, D. T., Floden, D., Alexander, M. P., Levine, B., & Katz, D. (2001). Stroop performance in focal lesion patients: dissociation of processes and frontal lobe lesion location. *Neuropsychologia*, 39(8), 771–786. [https://doi.org/10.1016/S0028-3932\(01\)00013-6](https://doi.org/10.1016/S0028-3932(01)00013-6)
- Sumiya, H., & Healy, A. F. (2004). Phonology in the bilingual Stroop effect. *Memory & Cognition*, 32(5), 752–758. <https://doi.org/10.3758/BF03195865>
- Tilsen, S. (2009). Subphonemic and cross-phonemic priming in vowel shadowing: Evidence for the involvement of exemplars in production. *Journal of Phonetics*, 37(3), 276–296. <https://doi.org/10.1016/j.wocn.2009.03.004>
- Tilsen, S. (2013). Inhibitory mechanisms in speech planning maintain and maximize. In A. Yu (Ed.), *Origins of sound*

- change: Approaches to phonologization* (pp. 112–117). Oxford University Press.
- Tipper, S. P., Howard, L. A., & Jackson, S. R. (1997). Selective reaching to grasp: Evidence for distractor interference effects. *Visual Cognition*, 4(1), 1–38. <https://doi.org/10.1080/713756749>
- Tourville, J. A., Reilly, K. J., & Guenther, F. H. (2008). Neural mechanisms underlying auditory feedback control of speech. *Neuroimage*, 39(3), 1429–1443. <https://doi.org/10.1016/j.neuroimage.2007.09.054>
- Van der Stigchel, S., & Theeuwes, J. (2008). Differences in distractor-induced deviation between horizontal and vertical saccade trajectories. *Neuroreport*, 19(2), 251–254. <https://doi.org/10.1097/WNR.0b013e3282f49b3f>
- Van Segbroeck, M., Travadi, R., Vaz, C., Kim, J., Black, M. P., Potamianos, A., & Narayanan, S. S. (2014, September). Classification of cognitive load from speech using an i-vector framework. *Interspeech*, 2014.
- Wedel, A., Kaplan, A., & Jackson, S. (2013). High functional load inhibits phonological contrast loss: A corpus study. *Cognition*, 128(2), 179–186. <https://doi.org/10.1016/j.cognition.2013.03.002>
- Wedel, A., Nelson, N., & Sharp, R. (2018). The phonetic specificity of contrastive hyperarticulation in natural speech. *Journal of Memory and Language*, 100, 61–88. <https://doi.org/10.1016/j.jml.2018.01.001>
- Wheeldon, L. (2003). Inhibitory form priming of spoken word production. *Language and Cognitive Processes*, 18(1), 81–109. <https://doi.org/10.1080/01690960143000470>
- Wiener, D., Tabor Connor, L., & Obler, L. (2004). Inhibition and auditory comprehension in Wernicke's aphasia. *Aphasiology*, 18(5-7), 599–609. <https://doi.org/10.1080/02687030444000228>
- Wittels, P., Johannes, B., Enne, R., Kirsch, K., & Gunga, H. C. (2002). Voice monitoring to measure emotional load during short-term stress. *European Journal of Applied Physiology*, 87(3), 278–282. <https://doi.org/10.1007/s00421-002-0625-1>
- Wright, R. (2004). Factors of lexical competition in vowel articulation. In J. Local, R. Ogden, & R. Temple (Eds.), *Papers in laboratory phonology VI* (pp. 26–50). Cambridge University Press.
- Yaniv, I., Meyer, D. E., Gordon, P. C., Huff, C. A., & Sevald, C. A. (1990). Vowel similarity, connectionist models, and syllable structure in motor programming of speech. *Journal of Memory and Language*, 29(1), 1–26. [https://doi.org/10.1016/0749-596X\(90\)90007-M](https://doi.org/10.1016/0749-596X(90)90007-M)