Journal of Memory and Language 143 (2025) 104649

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

Tomato-tomahto: Phonological representations vs. surface-level features in speech planning

Brett R. Myers [a,*], Cassandra L. Jacobs [b], Andrés Buxó-Lugo [c], Duane G. Watson [d]

[a] Department of Communication Sciences and Disorders, University of Utah, Salt Lake City, UT, United States
[b] Department of Linguistics, University at Buffalo, SUNY, Buffalo, NY, United States
[c] Department of Psychology, University at Buffalo, SUNY, Buffalo, NY, United States
[d] Department of Psychology and Human Development, Vanderbilt University, Nashville, TN, United States

ARTICLE INFO

Keywords:
Psycholinguistics
Speech production
Speech planning
Phonological encoding

ABSTRACT

Speakers often lengthen the duration of a word when it shares initial phonological segments with a previously uttered word (e.g., *candy* and *candle*). One explanation for this is that words with initial similarity affect phonological encoding during sequence planning, yet it is unclear whether this similarity is phonetic or phonological. We manipulated phonetic differences by using a dialect variant: the *pin-pen* merger in American English. Participants completed an event description task in three experiments. We manipulated whether the participant's target vowel ([ɪ] or [ɛ]) either phonetically matched or mismatched the vowel of the prime speaker, depending on the participant's dialect. In the second experiment, we introduced a control vowel in the prime word ([æ] vs. [ɛ]). Participants in both dialect groups lengthened target words when they shared an initial phoneme, even when the vowel of the overlapping prime word was not shared across dialects. In the third experiment, we replicated this finding in a larger cohort of non-merger participants. All three experiments showed word lengthening despite the phonetic realization of phonemes, suggesting this effect is driven by phonological representations rather than surface-level pronunciations.

Introduction

In spoken language, after a speaker has chosen the words they want to say and how those words should be combined, they must convert those words into sequences of sounds so that a listener can receive the message. The process of selecting the appropriate sounds and arranging them in the appropriate order is called "phonological encoding." Models of phonological encoding in language production typically agree that speakers access phonological content via the lexical representations of the planned message; however, many details about this selection process remain unknown, and studies suggest that the process is unlikely to be as simple as it may first appear. For example, numerous studies have shown that the phonological content of English speakers' recent productions can affect the duration of subsequent words if they share sounds. When words are repeated from a recent utterance, the second iterations often have reduced durations (e.g., Fowler & Housum, 1987; Lam & Watson, 2010; Buxó-Lugo, Toscano, & Watson, 2020; Jacobs, Yiu, Watson, & Dell, 2015). Conversely, producing words with *partial* overlap in their initial segments (e.g., *candy* and *candle*) in succession results in longer word durations for the second word (Myers & Watson, 2019; Watson, Buxo Lugo, & Simmons, 2015; Yiu & Watson, 2015).

The tendency to slow down when producing two words with initial segmental overlap was first demonstrated by Sevald and Dell (1994), who asked participants to repeat two-word phrases as quickly as possible. They found that producing words with initial overlap (e.g., *pick-pin*) yielded slower speech rates than producing words with final (rhyme) overlap (e.g., *pick-tick*). They explained this result using a model in which phonemes are accessed in sequential order, and activation of one phoneme triggers activation of the next phoneme. When word onsets overlap (*pi-*), both words in the pair become activated (*pick* and *pin*), resulting in competition that the production system must resolve to produce the intended next phoneme (–k or –n) in the current word. Similar results to Sevald and Dell have been found more recently in more naturalistic description tasks with effects on phonetic duration (Yiu & Watson, 2015; Watson et al., 2015; Buxó-Lugo, Jacobs, & Watson, 2020; Myers & Watson, 2019).

Interestingly, slowdowns of speech rate can occur independent of whether the speaker or someone else produces the first word in an

* Corresponding author at: 417 Wakara Way, Salt Lake City, UT, United States.
E-mail address: brett.myers@hsc.utah.edu (B.R. Myers).

https://doi.org/10.1016/j.jml.2025.104649
Received 18 March 2024; Received in revised form 27 April 2025; Accepted 28 April 2025
Available online 30 April 2025

overlapping pair. Buxó-Lugo, Jacobs, and Watson (2020) used an event description task in which participants named images on a screen that would shrink or flash. Participants either produced or heard a recording of the first word (the *prime*), or silently mouthed the prime, or did not mention it. The participants always said the second word (the *target*) out loud. They found that initial phonological overlap led to longer target word duration when the prime word was said aloud—either by the participant or someone else, but there was no lengthening when the prime word was silently mouthed or unmentioned. This suggests that the speaker's auditory memory of the prime word is a critical anchor point for target productions. That is, simply hearing a word can trigger slowdowns when producing a similar-sounding word, while silent articulation does not. Jacobs, Yiu, Watson, and Dell (2015) showed a similar effect using repetition reduction, in which repeated words and homophones were reduced when said aloud but not when unnamed or silently mouthed. These findings demonstrate the importance of the auditory experience of a prime word in the temporal dynamics of phonological encoding.

An open question is understanding the nature of the representations that are selected in the serial ordering process. The serial selection and ordering of sounds may lead one to believe that speech sounds are organized like beads on a string, with each sound occurring as a discrete, invariant unit. However, we know that the same sound can vary depending on context, co-articulation, and reduction (Shattuck-Hufnagel, 2019; Turk & Shattuck-Hufnagel, 2020). For instance, the final /n/ in *fun* may become /ng/ when one claims to be a *fun guy*. The final /s/ in *this* may be removed completely when referring to *this shirt*. Some words like "probably" may be reduced in a variety of ways ([ˈprɒbəbli], [ˈprɒbli], or [ˈprɒli]), and some groups of words may be reduced, as in "jever" (*did you ever*) or "io-noh" (*I don't know*). Despite surface-level differences, listeners are still able to extract the speaker's meaning (Niebuhr & Kohler, 2011; Manuel, 1995; Gow, 2001). This phonetic variation makes it clear that there are differences between how a sound is realized (phonetic representation) and its abstract category (phonological representation). The question we pose here aims to pinpoint at which level such encoding occurs. Said another way, does the ordering of sounds occur at an abstract phonological level or at the level of phonetic realization? To tackle this question, we leverage dialect differences in the realization of vowels in American English.

The pin-pen merger in Southern American English

We take advantage of a difference between varieties of English in the United States that do and do not have the *pin-pen* merger to test whether serial ordering is driven by phonological or phonetic representations. This dialect feature is a merger of the vowels /ɪ/ and /ɛ/ before nasal consonants across a wide geographical region, including most of the U.S. South, parts of the Midwest, and parts of California (Labov et al., 2006; Austen, 2020; Brown, 1991), as well as varieties of African-American English across the U.S. Traditionally, this merger has been described as merging toward [ɪ], making "pen" sound like "pin" (e.g., Brown, 1991; Bakos, 2013). However, Austen (2020) demonstrates that the merger may also favor [ɛ] in production, such that "twin" sounds like "twen.".

While most work on the *pin-pen* merger has focused on production, Austen (2020) found that individuals may also be merged in perception. She found a subset of Americans who cannot hear the difference between "pin" and "pen", even though their own productions are distinct. This is known as a near-merger, and it occurs about half as often as the full merger. Furthermore, Austen suggests that the merger may be reversing itself because she found it more prevalent in older rather than younger individuals. Still, the *pin-pen* merger is a natural form of phonetic variation that is common across the United States.

To better understand the type of sound sequence representations that are used in speech planning, we designed three experiments examining word lengthening effects by manipulating whether participants heard a speaker with their own dialect or one with the opposite pattern with respect to the *pin-pen* merger. The logic of the experiments below is as follows: speakers hear a recorded prime and then produce a target utterance. The prime and target words overlap in their initial sounds, so we expect lengthening of the target word. Importantly, the dialect in the recorded prime either matched or mismatched the dialect of the participant. We have seen that words lengthen when someone hears or utters a phonologically overlapping word to one that they are about to produce (Buxó-Lugo et al., 2020). If the acoustic details of the prime dialect meaningfully differ from that of the participant, we expect this to modulate the effect of phonological overlap. That is, if the serial ordering of phonemes relies on phonetic cues, then the degree of lengthening should be increased when the dialects are shared because their pronunciations will be more similar to the primes. However, if serial ordering happens at an abstract level beyond the acoustic realization of words, then we expect the effect of overlap on phonetic duration will be comparable regardless of whether the participant and prime dialects match or mismatch, as the underlying representations will be effectively identical.

Data availability

The data and materials collected and analyzed during the current study are available in the Open Science Framework repository, https://osf.io/pg5nz/?view_only=0916637535e34e5e8001db5c247348c0.

Experiment 1

Experiment 1 tests whether participants are sensitive to differences in the phonetic and phonological forms of the words they hear. We presented prime-target pairs that shared phonological segments while manipulating the phonetic detail of the prime (i.e., including the *pin-pen* merger or not).

Methods

Participants

There were 52 participants (age range: 18–22 years old, $M = 19.6$, $SD = 1.3$, 36 female) in Experiment 1. Participants were recruited from the Vanderbilt University Psychology Department subject pool and received course credit. All participants provided informed consent in accordance with the Vanderbilt University Institutional Review Board (IRB #160070).

Dialect screening

At the start of the study visit, participants completed a brief screening to determine whether their dialect included the *pin-pen* merger or not. For this screening, they were instructed to read aloud a paragraph presented on a computer screen. The paragraph was designed for this study, and it included eleven merger words (see Appendix). Participants were recorded via a head-mounted microphone at a sampling rate of 44,100 Hz. The first author classified speakers as having the merger or not based on auditory perception of the recorded passage.

There were 42 participants who spoke with the non-merged dialect and 10 participants were identified as having the *pin-pen* merger by perceptual assessment. This assessment was supported by analysis of vowel formants in the screening passage. The passage included 11 instances of merger-specific [ɛ] and 11 instances of [ɪ]. The Montreal Forced Aligner (McAuliffe et al., 2017) identified time points of these vowels, and a Praat script captured the F1 and F2 formants at the midpoint of each vowel. The mean formant values across all participants are shown in Fig. 1.

All participants were randomly assigned to hear a prime speaker who either had the merger or did not. The prime speakers were two females—one who speaks with the *pin-pen* merger and one who does not. The

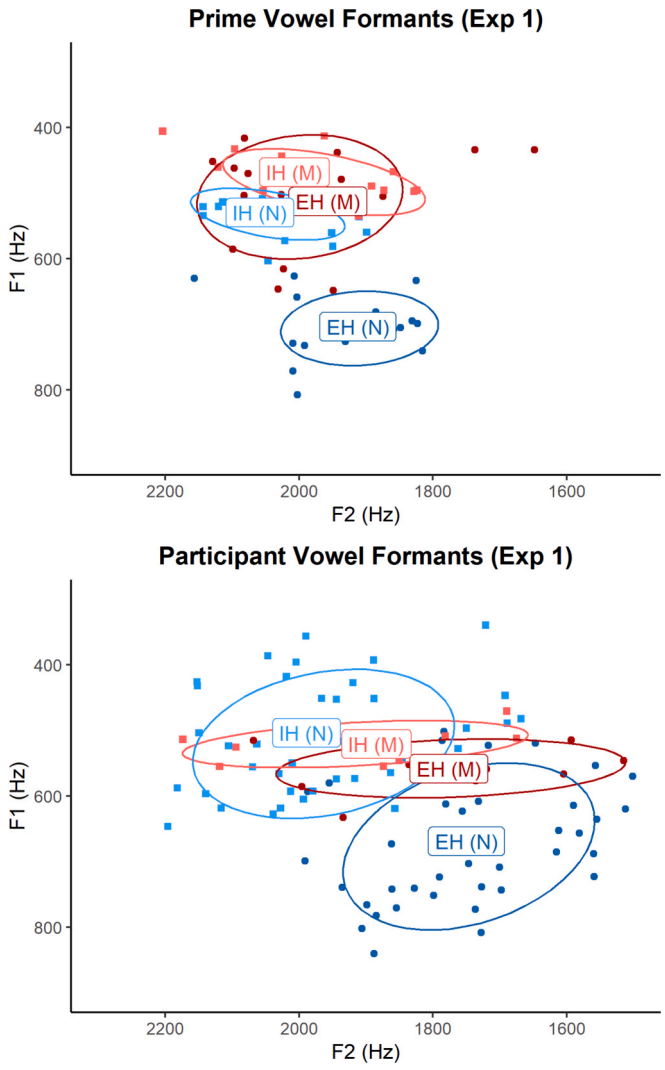

Fig. 1. F1-F2 formant values for [ɪ] ("IH") and [ɛ] ("EH") in Experiment 1. Top: The prime word vowels are plotted for both prime speakers—merger ("M") and non-merger ("N"). Bottom: Each participant's mean F1-F2 formants for the screening words are plotted. Non-merger ("N") speakers show a separation between "IH" and "EH", whereas merger ("M") speakers show an overlap. The ellipses represent 68% confidence intervals or one standard deviation.

speakers were recorded in an audiometric sound booth, and they were asked to read the same list of stimuli sentences with comfortable pitch and loudness. The formant analysis used for participants was also applied to the prime speaker recordings. The prime vowel formant values are displayed in Fig. 1. By random assignment, 25 participants heard a prime speaker that matched their own dialect, and 27 participants heard a speaker that did not match their dialect. Of the participants with the merger, 5 heard the merged prime, and 5 heard the non-merged prime. Of the non-merged participants, 22 heard the merged prime, and 20 heard the non-merged prime.

The prime speaker vowels were compared to verify that participants were presented with distinct vowels for the non-merged speaker but overlapping [ɪ] and [ɛ] vowels for the merged speaker. The degree of overlap between F1 and F2 values of the vowels was calculated as a Pillai score. A Pillai score is an output of a MANOVA, and it is an approach to determine how much two distributions (i.e., vowel formants) overlap (Hay, Warren, & Drager, 2006; Nycz & Hall-Lew, 2014). Pillai values range from 0 to 1, where 0 indicates complete overlap and 1 indicates distinct distributions. While there is generally no established threshold to demonstrate that two vowels are merged (Stanley & Sneller, 2023), Austen (2020) found that in the *pin-pen* merger, Pillai scores below 0.4 are a good indicator that [ɪ] and [ɛ] are merged. We performed a MANOVA for each pair of vowels per speaker. Our data are consistent with a high level of overlap (Pillai = 0.016) in [ɪ] and [ɛ] for the merged speaker (see Table 1.).

Materials

A set of 141 color images was selected from the Snodgrass and Vanderwart (1980) dataset (Rossion & Pourtois, 2001) and Clipart-style web images. A subset of 54 images served as the critical items, and the remaining images were filler items. Critical items consisted of 18 targets and 36 primes. There were two conditions for prime-target pairs:

1. **Overlap**: The *pendant* shrinks. The *penny* flashes.
2. **Control**: The *rabbit* shrinks. The *penny* flashes.

In the overlap condition, the prime-target pairs had phonological overlap in the initial segment. In the control condition, the prime-target pairs had no phonological overlap.

A Latin square design yielded four counterbalanced lists of items. Each participant was presented with 18 critical prime-target pairs, and each list had nine critical pairs for each of the two conditions. In addition, participants were exposed to 18 non-critical pairs, drawn from the filler items, for a total of 36 trials in the experiment. Trials were randomized for each participant.

Procedure

Participants completed the experiment on a Mac computer in MATLAB using the CogToolbox (Fraundorf et al., 2014) and Psychophysics Toolbox 3 (Kleiner et al., 2007). Following the dialect screening, participants completed a training task to learn the names of potentially ambiguous items. Items were displayed in the center of the screen with the intended label at the top of the screen, and participants recited the label aloud. They were reminded to use these names during the experiment.

Participants were then instructed on the experimental task. For each trial, four images were displayed equidistant around the center of the screen (see Fig. 2). The prime image would shrink, and an audio recording described the action ("The [image] shrinks"). As described above, participants heard either the merger or non-merger prime speaker for the duration of the experiment. Then the target image would flash, and participants described the action ("The [image] flashes"). Events occurred in the same order for all trials (i.e., shrinking then flashing). Trials were randomized and separated into two blocks, allowing participants to take a break between blocks as needed.

Acoustic analysis

Participant responses were segmented manually in Praat (Boersma & Weenink, 2017) to measure the duration of the target words. Three coders analyzed a subset of all trials in isolation using spectrographic and waveform information, and coders were blind to experimental condition of the trials. Coders were trained to identify target phonemes from spectrographic characteristics. Target words were segmented from start to end such that they were not identifiable as anything other than the targets (see Fig. 3). Duration of each target word was calculated with a custom Praat script.

Table 1
Pillai scores and *p*-values from MANOVA output for the [ɪ] and [ɛ] vowels for both prime speakers in Experiment 1.

	Non-merged Speaker		Merged Speaker	
Vowels	Pillai	*p*	Pillai	*p*
[ɪ] vs [ɛ]	0.775	< 0.001	0.016*	0.885

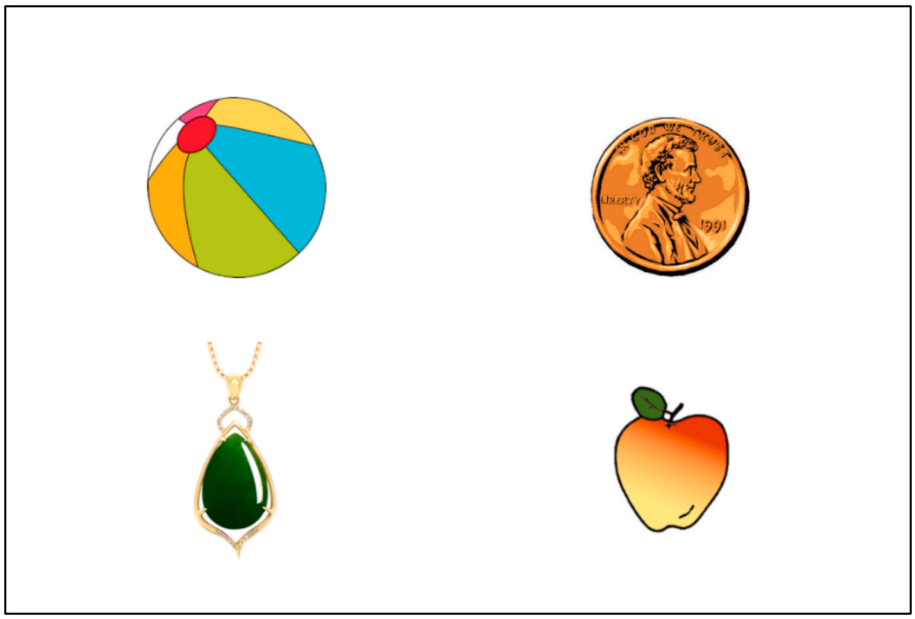

Fig. 2. Example layout of the experiment display consisting of four images in each quadrant of the screen. In this example, the *pendant* (prime) will flash and the *penny* (target) will shrink. The *ball* and *apple* are filler items.

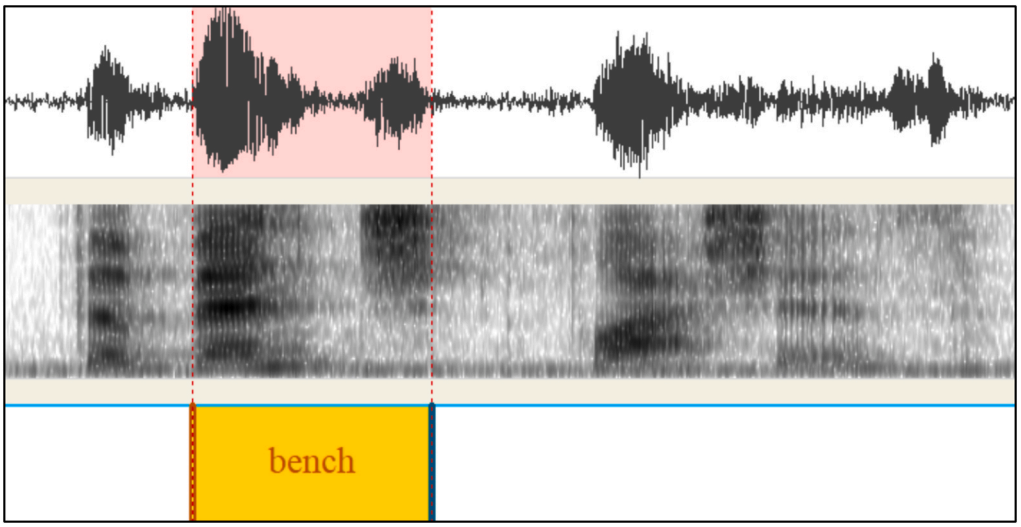

Fig. 3. Example word segmentation in Praat. Boundaries are placed at the start and end points of the target word: *bench*.

Statistical analysis

Following prior work with similar manipulations (Jacobs et al., 2020), target word durations were log transformed and analyzed using Bayesian linear mixed effects models with non-informative priors. We note that while there is some evidence that log transformation of durations and reaction times can lead to adverse inferences (Burchill & Jaeger, 2024), we report only the log transformed durations as the results were effectively identical when raw durations were used. Condition (sum coded with Overlap as 1 and No Overlap as –1), Dialect Match between the prime and participant (sum coded with Match as 1 and No Match as –1), and their interaction were included as fixed effects. Random slopes and intercepts by item and by participant were included for each main effect. Models were built using R package *brms* version 2.18.0 (Bürkner, 2017). We point out results where the 95 % credible interval for an estimated coefficient excludes 0.

Results

Target word durations were analyzed, and only target utterances that were the intended label were included. Trials were excluded if participants mispronounced the target, or if they used alternate labels (e.g., *doctor* for *dentist*, *chicken* for *hen*, or *coin* for *penny*). A total of 75 out of 936 trials met these exclusion criteria, leaving a total of 861 trials for analysis.

Our analysis finds evidence for an effect of condition, such that target words that had phonological overlap with the prime word were produced with longer durations than non-overlapping targets (see Table 2). Neither Dialect Match status nor the interaction between Condition and Dialect Match had 95 % credible intervals (CIs) that excluded 0; however, these effects were nearly observed. Fig. 4 shows average duration data by condition, prime dialect, and participant dialect.

Table 2
Bayesian multilevel model fixed effect estimates for log target word durations in Experiment 1. Effects with 95% CIs that exclude 0 are noted with an asterisk.

Fixed Effects	*β*	*Est. Error*	*95 % CIs*
Condition*	0.04	0.01	[0.03, 0.06]
Dialect Match	−0.03	0.02	[-0.08, 0.01]
Condition x Dialect Match	0.00	0.01	[-0.01, 0.01]

Discussion

In this experiment, we measured the duration of target words that either did or did not have phonological overlap with a prime. As predicted, there was evidence for lengthening in words with initial phonological overlap, which is consistent with previous studies investigating this effect (Yiu & Watson, 2015; Watson et al., 2015; Buxó-Lugo et al., 2020; Myers & Watson, 2019). This lengthening supports the notion that phonological representations can be primed and overlapping word forms can slow down production processes.

Our primary research question was understanding whether the acoustic realization of the prime phonemes matters in the overlap lengthening process. If phonological encoding is based on acoustic cues of phonemes, the lengthening effect should be modulated by the similarity between the dialects of the speaker and the prime. Contrary to this proposal, we see no evidence of an interaction between Dialect Match and Condition. That is, lengthening occurred regardless of whether the participant had the same dialect as the prime speaker or not. This provides initial evidence that phonological encoding is based on abstract representations of phonemes rather than the acoustic feature cues of phonemes.

Of course, an alternative explanation is that the lengthening effects are driven by the initial phoneme alone. That is, all primes minimally overlapped with targets in the first segment, in addition to varying in the vowel. To rule out this possibility, we conducted Experiment 2, which investigates the role of the vowel in these prime-target word pairs by including a condition in which the initial consonant and vowel overlap depending on dialect (overlap condition), a condition in which only the initial consonant overlaps (control vowel condition), and a condition with no overlap (control condition). If the effect of Experiment 1 was a result of consonant overlap alone, then we would expect the overlap and control vowel conditions to differ from the control condition but not from each other. On the other hand, if the effects were due to an abstract vowel representation, we would expect the overlap condition to be longer than the control vowel condition independent of dialect match.

Experiment 2

Methods

Participants

In Experiment 2, there were 48 participants (range: 18–22 years old, $M = 19.5$, $SD = 1.4$, 32 female). Participants were recruited and informed consent was obtained as in Experiment 1. Prior to analysis, three participants were excluded for not adhering to the instructions of the experiment (e.g., saying both the prime and target words).

Dialect screening

At the start of the study visit, participants completed the same screening as before to determine the presence of the *pin-pen* merger in their dialect. We used the same procedure from Experiment 1, using perceptual assessment and acoustic analysis to identify the merger participants. In this experiment, 37 participants had a non-merged dialect, and 8 participants were identified as having the *pin-pen* merger. The mean F1 and F2 formant values across all participants are shown in Fig. 5.

Participants were randomly assigned to hear a prime speaker who either had the merger or did not. This experiment used two different female speakers, and their mean vowel formants are displayed in Fig. 5. After random assignment, 20 participants heard primes that matched their own dialect, and 25 participants heard primes that did not match their dialect. Of the participants with the merger, 3 heard the merged prime, and 5 heard the non-merged prime. Of the non-merged participants, 20 heard the merged prime, and 17 heard the non-merged prime.

The prime speaker vowels were measured to verify that participants were presented with distinct vowels for the non-merged speaker but overlapping [ɪ] and [ɛ] vowels for the merged speaker. The degree of overlap between vowels was calculated as a Pillai score. We performed a MANOVA for each pair of vowels per speaker, which included the F1 and F2 values for all prime word vowels. Our data are consistent with a high level of overlap (Pillai = 0.018) in [ɪ] and [ɛ] for the merged speaker (see Table 3.).

Materials

A set of 222 color images was selected from the Snodgrass and Vanderwart (1980) dataset (Rossion & Pourtois, 2001) and Clipart. A

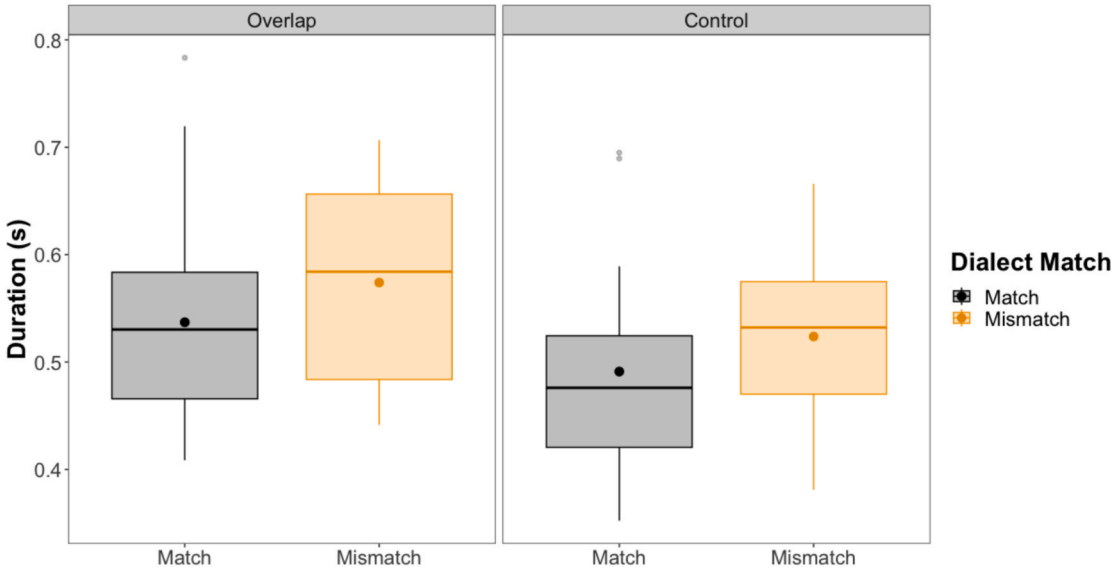

Fig. 4. Boxplots showing target word durations in seconds. Mean durations are represented by solid points.

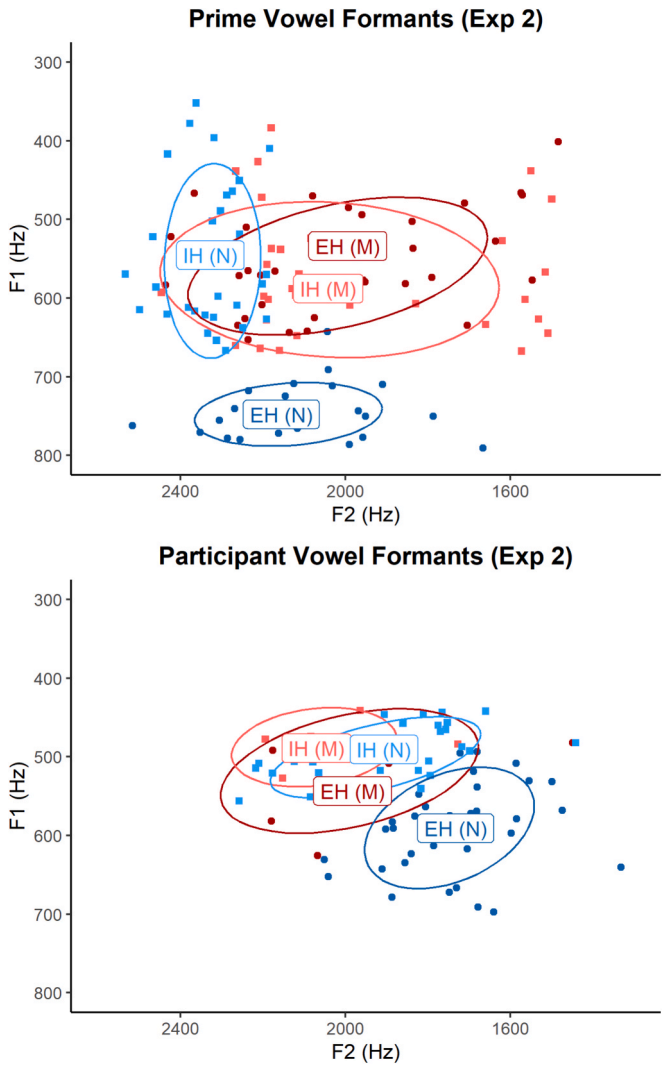

Fig. 5. F1-F2 formant values for [ɪ] ("IH") and [ɛ] ("EH") in Experiment 2. Top: The prime word vowels are plotted for both prime speakers—merger ("M") and non-merger ("N"). Bottom: Each participant's mean values for the screening words are plotted. The ellipses represent 68% confidence intervals or one standard deviation.

Table 3
Pillai scores and *p*-values from MANOVA output for the [ɪ] and [ɛ] vowels for both prime speakers in Experiment 2.

	Non-merged Speaker		Merged Speaker	
Vowels	Pillai	*p*	Pillai	*p*
[ɪ] vs [ɛ]	0.787	< 0.001	0.018*	0.759
[ɪ] vs [æ]	0.806	< 0.001	0.751	< 0.001
[ɛ] vs [æ]	0.742	< 0.001	0.455	0.002

subset of 132 images served as the critical items, and the remaining images were filler items. Critical items consisted of 33 targets, with 3 potential primes (i.e., 99 prime images in total) belonging to three conditions:

1. **Overlap**: The *pendant* shrinks. The *penny* flashes.
2. **Control Vowel**: The *panda* shrinks. The *penny* flashes.
3. **Control**: The *rabbit* shrinks. The *penny* flashes.

In the overlap condition, the prime-target pairs had phonological overlap as in Experiment 1. In the control vowel condition, the prime word had /æ/ in the initial syllable, while the target word had /ɛ/ in the same position. In the control condition, the prime-target pairs had no phonological overlap.

A Latin square design yielded six counterbalanced lists of items. Each participant was presented with 33 critical prime-target pairs, with eleven critical pairs for each prime condition. In addition, participants were exposed to 47 non-critical pairs, drawn from the filler items, for a total of 80 trials in the experiment. Trial order was randomized for each participant.

Procedure

This experiment used the same procedures as Experiment 1, including the training session at the beginning to learn target labels for ambiguous images. For each experimental trial, the prime image would shrink, and an audio recording described the action. Again, two different female talkers were used for the primes—one with the *pin-pen* merger and one without it. After hearing the audio recording, the target image would flash, and participants described the action. Trials were randomized and separated into four blocks, allowing for breaks as needed.

Statistical analysis

Participant responses were segmented as in Experiment 1 by three coders. The duration of each target word was calculated from a custom Praat script. Log transformed target word durations were analyzed using Bayesian mixed effects models with non-informative priors. Condition (dummy coded with the Control-vowel condition as the comparison group), Dialect Match (sum coded with Match as 1 and No-Match as –1), and their interactions were included as fixed effects, along with random slopes and intercepts by item and by participant for each of the effects except the interaction. We point out results where the 95 % credible interval for an estimated coefficient excludes 0.

Results

Only target utterances containing the intended word label were included in the analyses. A total of 400 out of 1485 trials were removed due to misnamed targets or a missing response, leaving a total of 1085 trials for analysis. Excluded trials were evenly distributed across conditions.

The models show evidence of an effect of Condition, such that target words with the control vowel were longer than non-overlapping targets, but also slightly shorter than target words in the overlap condition (see Table 4). Similar to results from Experiment 1, there was no evidence of an interaction between Condition and Dialect Match, suggesting that lengthening was not modulated by the phonetic similarity between productions from the speaker of the target word and the speaker of the prime. Fig. 6 shows average duration data by condition, prime dialect, and participant dialect.

Discussion

Experiment 2 examined whether the lengthening effects of overlap in Experiment 1 were due to competing abstract phonological encoding processes driven by shared initial consonants and vowels for targets and primes, or whether they were due to the target and prime simply sharing

Table 4
Bayesian multilevel model fixed effect estimates for log target word durations in Experiment 2. Effects with 95% CIs that exclude 0 are noted with an asterisk.

Fixed Effects	***β***	***Est. Error***	***95 % CIs***
Condition: Control vs. Control Vowel*	−0.13	0.02	[-0.17, −0.09]
Condition: Overlap vs. Control Vowel*	0.03	0.01	[0.01, 0.06]
Dialect Match	0.00	0.02	[-0.04, 0.05]
Dialect Match x Control	0.00	0.02	[-0.04, 0.04]
Dialect Match x Overlap	0.01	0.01	[-0.02, 0.03]

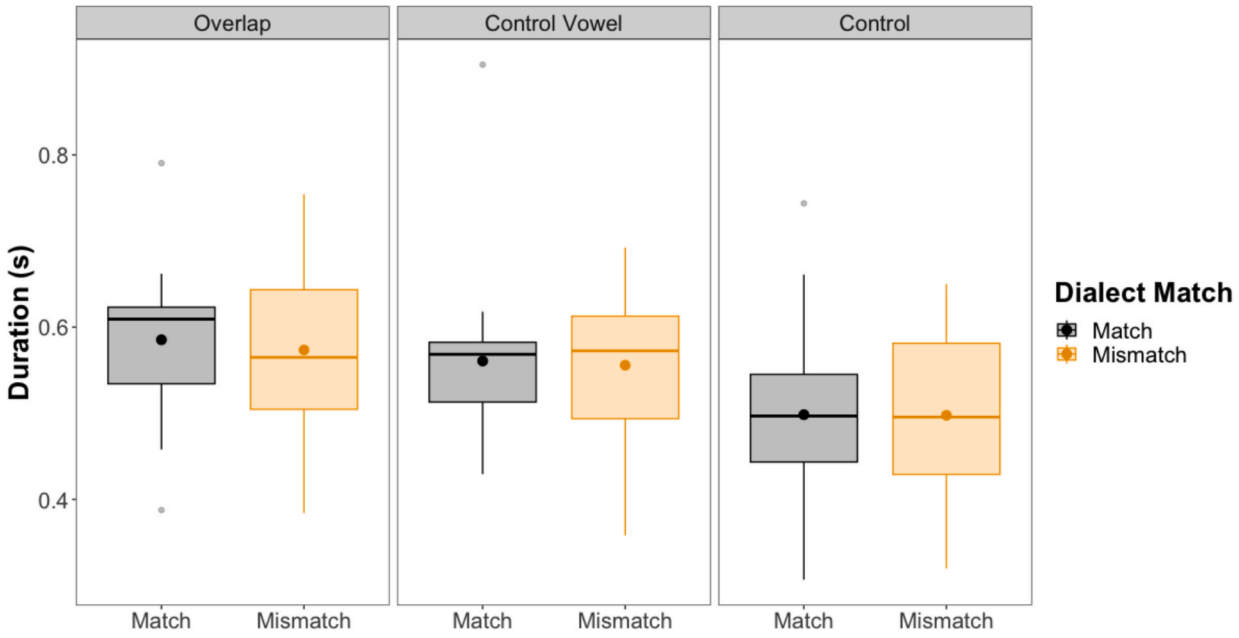

Fig. 6. Boxplots showing target word durations in seconds in Experiment 2. Mean durations are represented by solid points.

an initial consonant. The data from Experiment 2 suggest that both of these sources of overlap affected word durations.

Word pairs with initial consonant and vowel overlap (e.g., *pendant-penny*) showed the greatest lengthening. Word pairs with the control vowel /æ/ (e.g., *panda-penny*) also showed significant lengthening, suggesting the initial consonant is enough to elicit the lengthening effect. It is also possible that the overlapping nasalization from the post-vocalic consonant could have contributed to the lengthening effect. Critically, there was a significant difference between the overlap and control vowel condition, demonstrating that lengthening in the overlap condition occurred whether a participant's vowel production matched or did not match the prime speaker's production. There was also no significant Dialect Match by Condition interaction, suggesting that the treatment of the vowel conditions was not dependent on the prime dialect. Although this lack of an interaction is a null result, our Bayesian approach to the analysis suggests support for a lack of a difference between groups.

Experiment 3

One limitation in Experiment 2 is the relatively small number of observations to detect an interaction between Dialect Match and Condition. There may not have been enough participants to support the claim that word lengthening in the overlap condition happens regardless of the prime dialect. Therefore, we conducted a third experiment with the goal of recruiting more participants to answer the question—does the prime dialect matter in the overlap lengthening effect? To help clarify the results, we elected to include only participants with the non-merger dialect in the third experiment.

Methods

Participants

In the current experiment, 73 participants were recruited (range: 18–60 years old, $M = 28.8$, $SD = 9.73$, 66 female). Participants were volunteers recruited from campus announcements at the University of Utah. All participants provided informed consent in accordance with the University of Utah Institutional Review Board (IRB #00171949).

Dialect screening

At the start of the study visit, participants completed the same screening as Experiments 1 and 2 to determine the presence of the *pin-pen* merger in their dialect. We used the same procedure of perceptual assessment and acoustic analysis to identify the merger participants. In this experiment, we chose to only use data from individuals who had the non-merger dialect. Two participants exhibited the merger, and those two were excluded from data analysis. The remaining 71 participants were included in the study. Prime and participant formant values are displayed in Fig. 7.

Materials & procedures

The stimulus images and prime speakers used in this experiment were identical to Experiment 2. The same three conditions were used—overlap, control vowel, and control. As before, there were 33 critical prime-target pairs and 47 non-critical pairs, for a total of 80 trials. The same training session occurred at the beginning of the experiment to teach participants the target labels of potentially ambiguous images.

For each experimental trial, the prime image would shrink, and an audio recording described the action. Participants were randomly assigned to hear the speaker with the *pin-pen* merger or the one without the merger. After hearing the audio recording, the target image flashed, and participants described the action. Trials were randomized and separated into four blocks, allowing for breaks as needed.

Statistical analysis

Participant responses were segmented by five coders. The duration of each target word was calculated from a custom Praat script. Log transformed target word durations were analyzed using Bayesian mixed effects models with non-informative priors. Condition (dummy coded with the Control-vowel condition as the comparison group), Prime Dialect (sum coded with Merger as 1 and Non-Merger as –1), and their interactions were included as fixed effects, along with random slopes and intercepts by item and by participant for each of the effects except the interaction. Results were considered significant where the 95 % credible interval for an estimated coefficient excluded 0.

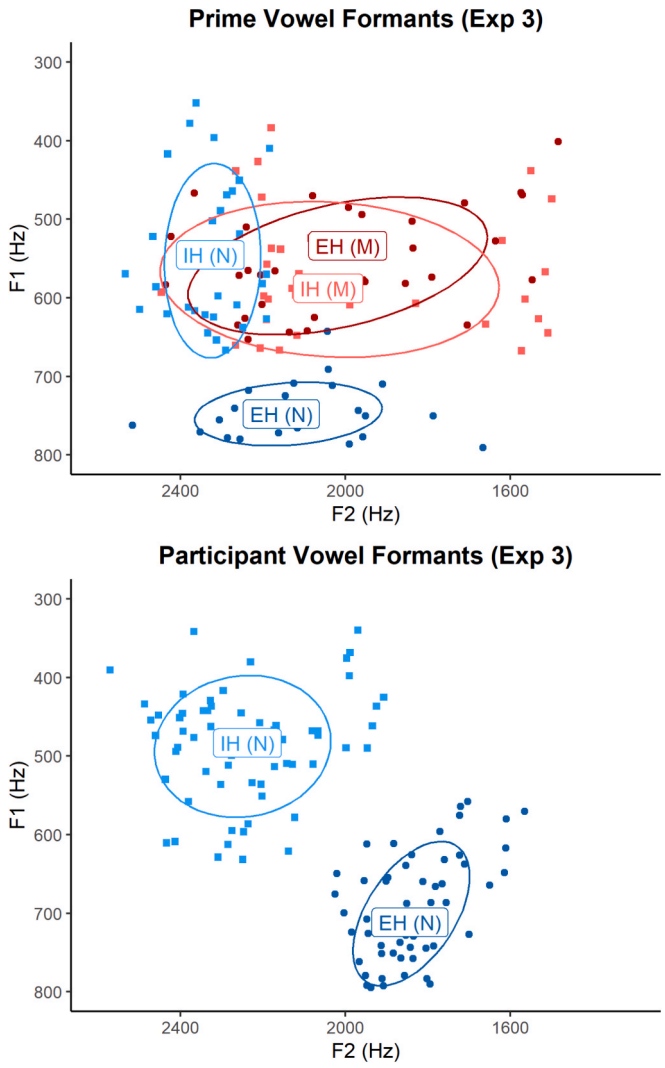

Fig. 7. F1-F2 formant values for [ɪ] ("IH") and [ɛ] ("EH") in Experiment 3. Top: The prime word vowels are plotted for both prime speakers—merger ("M") and non-merger ("N"). Note these are the same prime speakers as in Experiment 2. Bottom: Each participant's mean values for the screening words are plotted. The ellipses represent 68% confidence intervals or one standard deviation.

Results

Experiment 3 included 71 participants who were randomly assigned to a prime condition; 35 participants heard the merged prime speaker, and 36 participants heard the non-merged prime speaker. Out of 2,343 possible responses, 339 trials were removed due to misnamed targets or a missing response—leaving 2,004 trials for analysis. Excluded trials were evenly distributed across conditions.

The models show evidence of an effect of Condition, such that target words with the control vowel were longer than non-overlapping targets and also shorter than target words in the overlap condition (see Table 5).

Table 5
Bayesian multilevel model fixed effect estimates for log target word durations in Experiment 3. Effects with 95% CIs that exclude 0 are noted with an asterisk.

Fixed Effects	*B*	*Est. Error*	*95 % CIs*
Condition: Control vs. Control Vowel*	−0.08	0.01	[0.07, 0.09]
Condition: Overlap vs. Control Vowel*	0.04	0.00	[0.03, 0.05]
Prime Dialect*	0.03	0.01	[0.02, 0.04]
Prime Dialect x Control Vowel	0.01	0.01	[0.00, 0.02]
Prime Dialect x Overlap	0.00	0.00	[0.00, 0.01]

Similar to results from Experiments 1 and 2, there was no evidence of an interaction between Condition and Prime Dialect. This lack of an interaction effect is replicated now with more participants to support the claim. Fig. 8 shows average duration data by condition and prime dialect.

General discussion

This study aimed to examine the phonological overlap lengthening effect in the presence of dialect variation using the *pin-pen* merger. Specifically, we were interested in testing the hypothesis that phonological encoding could rely on acoustic properties of phonemes during speech production. In Experiment 1, we observed word lengthening in prime-target pairs with initial phonological overlap regardless of dialect differences. In Experiment 2, we replicated this finding and ruled out the possibility that the effect could be wholly explained by phonological overlap of only the initial consonants in prime-target pairs. In Experiment 3, we replicated the same findings with more participants to support the claim. The overlap condition had the longest word durations, and it did not matter whether the acoustic realization of the vowel was [ɛ] or [ɪ]. Overall, word lengthening occurred regardless of vowel variation in the stimulus, suggesting abstract representations affect phonological encoding (Fig. 9).

The primary question for these experiments was to determine if acoustic properties or abstract phonological representations are responsible for the word lengthening effect. Dialect variation was a means to present naturalistic acoustic differences among repeated phonemes. Future work may choose to replicate these experiments with different mergers or other dialect features. One example would be the glottal stop in the Cockney dialect of English. Researchers could compare an American pronunciation of "butter" with a medial tap /ɾ/ versus a Cockney glottal stop /ʔ/. This would present a clear acoustic distinction that is likely more obvious than the *pin-pen* separation used in this study. It may also be interesting to compare *pin-pen* stimuli with *pit-pet* stimuli to further test predictions of phonological versus phonetic priming. These alternatives would provide additional investigations into acoustic differences, and we leave that for future work. Still, the three experiments presented here consistently demonstrate that word lengthening is driven by phonological priming.

We acknowledge that dialect variation is complicated in the real world, and our participants likely have had exposure to the *pin-pen* merger whether they use it or not. We do not make claims about how one's dialect may influence the lengthening effect or how this effect will manifest in cross-dialect communication. Furthermore, we did not attempt to categorize or label the dialects of our participants; our primary interest was to determine the presence or absence of the *pin-pen* merger as an acoustic feature of their speech patterns. As such, we are able to make the basic claim about whether two speakers had an acoustic match or mismatch of the target segments, which allowed us to directly address our research question. Regardless of the phonetic properties of the vowel, listeners seem to map surface-level acoustics onto their own phonological representation of the vowel.

In Experiment 3, there was also a main effect of prime dialect. Participants who heard the merged prime produced significantly longer word durations than participants who heard the non-merged (standard) prime. This is likely attributed to the merged speaker having a slightly slower rate of speech than the non-merged speaker. This slower speech rate is a characteristic of speakers who use the merger (e.g., Clopper & Smiljanic, 2015). It is possible that participants aligned their own rate of speech to the prime speaker's rate. Previous work has shown that speakers converge to a shared rate of speech in conversational settings, and this rhythmic entrainment is mutually beneficial for the interlocutors (e.g., Manson et al., 2013; Freud et al., 2018; Cohen Priva et al., 2017). Additionally, speakers will converge to characteristics—such as speech rate—when listening to a non-native speaker (Wagner et al., 2021). Critically, we see no evidence that this difference

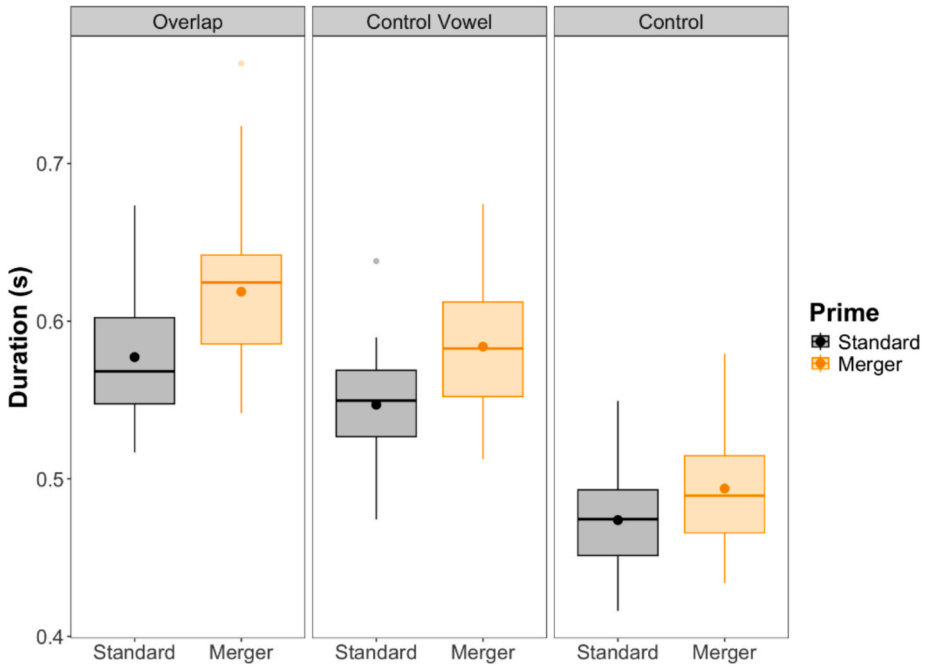

Fig. 8. Boxplots showing target word durations in seconds in Experiment 3. Mean durations are represented by solid points. All participants either heard the merged prime or the non-merged (standard) prime.

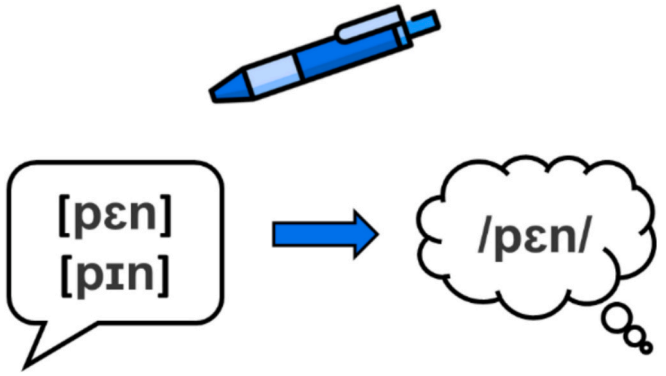

Fig. 9. Illustration of phonological encoding in the presence of dialect variation. Regardless of the stimulus dialect, the listener interprets the phonological sequence with an abstract representation of the segment.

in speech rate interacted with the lengthening effect.

An important note about these experiments is that the effects we report are small. Of course, this raises the question of whether these small effects are due to the processes described above being peripheral to language production (and consequently of little interest to language production researchers), or whether the processes are central to language production, but the measures we use to detect them are relatively noisy. We think the latter is likely. We are trying to detect serial processing effects by measuring word duration in a comprehension to production, priming/interference paradigm. This paradigm is rarely used in the production literature and not as well understood as the "usual suspects" of psycholinguistics (e.g., eye fixations, reaction times, or speech onset times). Future work will need to explore the limits of the paradigm to better estimate the effect sizes reported here. Having said that, although the effects are small, the theoretical predictions are clear, and the data reported above suggest that abstract phonological representations are used in serially ordering speech.

The critical finding in these experiments was that word lengthening occurred despite differences in the acoustic properties of speech. This supports the idea that speech planning operates over abstract phonological representations. In the instances where a participant dialect did not match the prime speaker's, we hypothesize that the participant hears the given phoneme and encodes their own representation of that phoneme. It is not evident that the participant encodes the acoustic properties of the phoneme in a way that interferes with their own production. Although this finding does not rule out the role of surface-level acoustic features in speech planning, it does suggest that abstract representations play a central role in speech production.

CRediT authorship contribution statement

Brett R. Myers: Writing – review & editing, Writing – original draft, Visualization, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Cassandra L. Jacobs:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Conceptualization. **Andrés Buxó-Lugo:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis. **Duane G. Watson:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by funding from the National Science Foundation [Grant Number 1557097]. Any opinions, findings, and conclusions, or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. The authors would like to thank Maya Ricketts, Ella Dixon, Jenna DiStefano, Russell deJesus, Darla Lowe, Emily Carter, and Danielle Toth for assistance with data collection and analysis.

Appendix A

Reading passage for pin-pen *merger screening. Eleven merger words are italicized.*

My *friend* and I *went* to an Italian restaurant to celebrate her birthday. She ordered the *penne* pasta and I asked for linguine. Our server *Wendy* took our *menus*, and before we knew it, *Ken* the *bartender* was pouring champagne for both of us. The manager—I think her name was Julie—brought us our food. *Then* we heard music and *ten* servers came over to sing Happy Birthday, and they placed a large chocolate cake in the *center* of the table. It was a birthday to *remember*.

Appendix B

Critical Stimuli for Experiment 1.

Prime	Target
bend	bench
center	sentence
den	dent
denim	dentist
fender	fencer
henge	hen
lentil	lender
membrane	member
memo	menu
pendant	penny
penguin	pencil
sender	centaur
sensor	senate
temper	temple
tempest	tempo
ten	tent
tendon	tender
trend	trench

Appendix C

Critical Stimuli for Experiments 2 & 3.

Overlap Prime	Control Vowel Prime	Target
bend	band	bench
center	sandwich	sentence
century	sandpaper	centerpiece
den	dance	dent
denim	dancer	dentist
embryo	amazon	emperor
employee	amplitude	empire
end	and	en
enemy	animal	energy
enzyme	antler	entrance
fender	phantom	fencer
gentleman	janitor	general
henge	hand	hen
lemming	lampshade	lemon
lens	land	length
lentil	landing	lender
membrane	mantle	member
memo	mango	menu
pendant	panther	penny
penguin	pancake	pencil
pentagon	panelist	pendulum
penthouse	panda	penlight
rent	ranch	wrench
sender	sander	centaur
sensor	sandbox	senate
stent	stand	stench
temper	tanker	temple
tempest	tango	tempo
ten	tam	tent

(*continued on next page*)

(*continued*)

Overlap Prime	Control Vowel Prime	Target
tenant	tangent	tenor
tendon	tandem	tender
trend	tram	trench
venom	vandal	vendor

Data availability

Data will be made available on request.

References

Austen, M. (2020). Production and perception of the pin-pen merger. *Journal of Linguistic Geography, 8*, 115–126.

Bakos, J. (2013). A comparison of the speech patterns and dialect attitudes of Oklahoma. Stillwater, OK: Oklahoma State University dissertation.

Boersma, P., & Weenink, D. (2017). Praat: doing phonetics by computer (Version 6.0.37).

Brown, V. (1991). Evolution of the merger of /ɪ/ and /ɛ/ before nasals in Tennessee. *American Speech, 66*(3), 303–315.

Bürkner, P. C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software, 80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01

Burchill, Z. J., & Jaeger, T. F. (2024). How reliable are standard reading time analyses? Hierarchical bootstrap reveals substantial power over-optimism and scale-dependent Type I error inflation. *Journal of Memory and Language, 136*, Article 104494.

Buxó-Lugo, A., Jacobs, C. L., & Watson, D. G. (2020). The world is not enough to explain lengthening of phonological competitors. *Journal of Memory and Language, 110*, Article 104066.

Clopper, C. G., & Smiljanic, R. (2015). Regional variation in temporal organization in American English. *Journal of Phonetics, 49*, 1–15.

Cohen Priva, U., Edelist, L., & Gleason, E. (2017). Converging to the baseline: Corpus evidence for convergence in speech rate to interlocutor's baseline. *J. Acoust. Soc.,, 141*(5).

Fowler, C. A., & Housum, J. (1987). Talkers' signaling of "new" and "old" words and listeners' perception and use of the distinction. *Journal of Memory and Language, 26*, 489–504.

Fraundorf, S.H., Diaz, M.I., Finley, J.R., Lewis, M.L., Tooley, K.M., Isaacs, A.M., Lam, T. Q.,Trude, A.M., Brown-Schmidt, S., & Brehm, L. (2014). CogToolbox for MATLAB [computer software].

Freud, D., Ezrati-Vinacour, R., & Amir, O. (2018). Speech rate adjustment of adults during conversation. *Journal of Fluency Disorders, 57*, 1–10.

Gow, D. (2001). Assimilation and anticipation in continuous spoken word recognition. *Journal of Memory and Language, 45*, 133–159.

Hay, J., Warren, P., & Drager, K. (2006). Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics, 34*, 458–484.

Jacobs, C. L., Loucks, T. M., Watson, D. G., & Dell, G. S. (2020). Masking auditory feedback does not eliminate repetition reduction. *Language, Cognition and Neuroscience, 35*(4), 485–497. https://doi.org/10.1080/23273798.2019.1693051

Jacobs, C. L., Yiu, L. K., Watson, D. G., & Dell, G. S. (2015). Why are repeated words produced with reduced durations? Evidence from inner speech and homophone production. *Journal of Memory and Language, 84*, 37–48.

Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in Psychtoolbox-3? *Perception, 36*(14), 1–16.

Labov, W., Ash, S., & Boberg, C. (2006). *The Atlas of North American English*. Berlin: Mouton de Gruyter.

Lam, T. Q., & Watson, D. G. (2010). Repetition is easy: Why repeated referents have reduced prominence. *Memory & Cognition, 38*, 1137–1146.

Manson, J. H., Bryant, G. A., Gervais, M. M., & Kline, M. A. (2013). Convergence of speech rate in conversation predicts cooperation. *Evolution and Human Behavior, 34*(6), 419–426.

Manuel, S. Y. (1995). Speakers nasalize /dh/ after /n/, but listeners still hear /dh/. *Journal of Phonetics, 23*(4), 453–476.

McAuliffe, M., Socolof, M., Stengel-Eskin, E., Mihuc, S., Wagner, M., & Sonderegger, M. (2017). Montreal Forced Aligner [Computer program]. *Version, 1*.

Myers, B., & Watson, D. (2019). Paying the meter: Effects of metrical similarity on word lengthening. *Psychonomic Bulletin & Review, 26*(6), 1941–1947.

Niebuhr, O., & Kohler, K. (2011). Perception of phonetic detail in the identification of highly reduced words. *Journal of Phonetics, 39*, 319–329.

Nycz, J., & Hall-Lew, L. (2014). Best practices in measuring vowel merger. In *Proceedings of Meetings on Acoustics* (p. 20).

Rossion, B., & Pourtois, G. (2001). Revisiting Snodgrass and Vanderwart's object database: Color and texture improve object recognition. *Journal of Vision, 1*, 413a.

Sevald, C. A., & Dell, G. S. (1994). The sequential cuing effect in speech production. *Cognition, 53*, 91–127.

Shattuck-Hufnagel, S. (2019). Toward an (even) more comprehensive model of speech production planning. *Language, Cognition, and Neuroscience, 34*(9), 1202–1213.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, familiarity and visual complexity. *Journal of Experimental Psychology: Human Learning & Memory, 6*, 174–215.

Stanley, J. A., & Sneller, B. (2023). Sample size matters in calculating Pillai scores. *The Journal of the Acoustical Society of America, 153*(1), 54–67.

Turk, A. E., & Shattuck-Hufnagel, S. (2020). *Speech timing*. Oxford: Oxford University Press.

Wagner, M. A., Broersma, M., McQueen, J. M., Dhaene, S., & Lemhöfer, K. (2021). Phonetic convergence to non-native speech: Acoustic and perceptual evidence. *Journal of Phonetics, 88*.

Watson, D. G., Buxo Lugo, A. S., & Simmons, D. C. (2015). *The Effect of Phonological Encoding on Word Duration: Selection Takes Time* (Vol. 46,, 85–98.

Yiu, L. K., & Watson, D. G. (2015). When overlap leads to competition: Effects of phonological encoding on word duration. *Psychonomic Bulletin & Review, 22*, 1701–1708.