HHS Public Access

Author manuscript

J Exp Psychol Gen. Author manuscript; available in PMC 2026 January 13.

Speech Motor Control is Not Sequestered from General Auditory Processes

Craig Thorburn[1], **Lin Zhou**[1], **Frederic Dick**[2], **Nazbanou Nozari**[3], **Lori L. Holt**[1]

[1] Department of Psychology, University of Texas at Austin

[2] Department of Experimental Psychology, University College London

[3] Department of Psychological and Brain Sciences, Indiana University

Abstract

There is growing recognition that short-term changes in speech perception influence speech production. These effects offer new insight into interactions of perception and production and shed light on phonetic convergence, the subtle alignment of speech patterns that emerges between communication partners. Across three experiments, we investigate the representations underlying perceptual effects on speech production. Building from the established influence of preceding context on speech perception, we strategically pair contexts to shift perception of target syllables and test whether these perceptual effects influence speech production. Experiment 1 shows that speech contexts rich in articulatory-phonetic information shift speech perception and alter acoustic patterns of speech production. Experiment 2 demonstrates that continuous natural speech filtered to possess subtly different spectral profiles that do not impact articulatory-phonetic information also affect both perception and production. Strikingly, Experiment 3 reveals that even nonspeech tones induce perceptual context effects that influence speech production. The findings point to a much broader scope of perception-production transfer than reported previously, and challenge the necessity of social interaction, covert imitation, and articulatory-phonetic information in sensorimotor speech interactions. This emphasizes the need to extend models of speech motor control to account for perceptual influences of other talkers' speech on speech production, and to accommodate general auditory processes in sensorimotor models of speech.

Keywords

Speech perception; speech production; phonetic convergence; spectral contrast; sensorimotor speech

Listening to another voice can influence one's own speech. For example, speakers subtly and unconsciously adjust their speech to sound more like each other in conversation

Correspondence should be addressed to LLH, lori.holt@austin.utexas.edu.

CRediT: CT: Data Curation, Formal Analysis, Investigation, Methodology, Software, Validation, Visualization, Writing - Original Draft, Writing - Review and Editing; **LZ:** Formal Analysis, Investigation, Methodology, Software, Validation, Visualization, Writing, Review and Editing; **FD:** Supervision, Writing - Review and Editing; **BN:** Writing - Review and Editing; **LLH:** Conceptualization, Funding Acquisition, Project Administration, Resources, Supervision, Validation, Visualization, Writing – Original Draft, Writing, Review and Editing

(Mukherjee et al., 2017; Murphy, Nozari & Holt, 2024; Pardo et al., 2018). This phonetic convergence is influenced by social factors such as likeability, social status, and attractiveness (Bourhis & Giles, 1977; Gregory & Webster, 1996; Michalsky & Schoormann, 2017), suggesting a role in fostering social connection (Giles et al., 1991, 2023). Yet, phonetic convergence also arises in non-interactive tasks (Murphy et al., 2023; Murphy, Nozari, & Holt, 2025; Pardo et al., 2018; Sato et al., 2013; Shockley et al., 2004), indicating that cognitive and perceptual mechanisms also contribute.

It has been difficult to predict the effects of others' speech on speech production. For instance, phonetic convergence emerges for some utterances and acoustic speech dimensions but not others (Earnshaw, 2021; Heath, 2015; Lindsay et al., 2022; Nielsen, 2011; Ostrand & Chodroff, 2021; Pardo et al., 2013; Schertz & Paquette-Smith, 2023). Sometimes convergence is greater among female speakers (Namy et al., 2002), but other times it is more pronounced in males (Pardo, 2006; Pardo et al., 2010) or it may exhibit more complex patterns (Miller et al., 2010; Pardo et al., 2017). This points to potentially complex and interacting contributions of perceptual, cognitive, social, and contextual factors (Babel, 2010; Bourhis & Giles, 1977; Giles et al., 1991; Pardo, 2006).

This complexity has motivated calls for a deeper understanding of the core perceptual and cognitive mechanisms linking speech perception and production (Babel, 2012; Pardo et al., 2022). Sensorimotor speech adaptation studied using altered auditory feedback designs provides insight into these core mechanisms (Bourguignon et al., 2016; Bradshaw et al., 2023; Lametti et al., 2014; Sato et al., 2013; Shiller & Rochon, 2014). In these studies, experimenters perturb the auditory feedback participants receive from their own speech, prompting adjustments to speech production (Houde & Jordan, 1998; Purcell & Munhall, 2006; Rochet-Capellan & Ostry, 2011; Tourville et al., 2008; Villacorta et al., 2007). Experimenters record and rapidly process the utterances to deliver an altered version back to participants with a very brief delay. For example, participants might utter words with the vowel /ɛ/ (as in *bet*) while auditory feedback is altered to lower first formant (F1) frequency, shifting vowel acoustics closer to /I/ as in *bit*. This altered auditory feedback results in compensatory adjustments to speech motor control such that participants subsequently produce /ɛ/ vowels with a *higher* F1 frequency (Bourguignon et al., 2016).

Sensorimotor adaptation driven by auditory feedback from one's own voice can be influenced by speech heard from *other voices* (Bourguignon et al., 2016; Lametti et al., 2014; Shiller & Rochon, 2014). Bourguignon and colleagues preceded each /ɛ/ utterance with speech contexts known to shift vowel perception (Ladefoged & Broadbent, 1957). Specifically, a carrier phrase (*Please say what this word is*) that has been filtered to exaggerate high frequency acoustic energy shifts vowel perception toward /I/ (with *lower* F1) whereas a phrase with *lower*-frequency energy emphasized shifts the same vowels to be more often reported as /ɛ/ (*higher* F1). Bourguignon et al. found that these perceptual context effects (evoked by listening to another voice) influence sensorimotor adaptation to altered auditory feedback from one's own voice. Namely, the compensatory adjustments to speech production elicited by lowering /ɛ/ F1 (toward /I/) are exaggerated when higher-frequency speech context from another voice pushes vowel *perception* toward /I/. Perception

of another talker's speech thus collaborates with perception of auditory feedback from one's own voice to impact speech motor control.

The perceptual influence of other talkers' speech on speech production is apparent in other contexts, as well. Explicit perceptual training that shifts listeners' vowel perceptual boundaries affects sensorimotor adaptation to altered auditory feedback (Lametti et al., 2014; Shiller & Rochon, 2014). Moreover, influences of perception on production are evident even without altered auditory feedback. For example, Murphy et al. (2024; 2025a,b) find that perceptual statistical learning shifts speech production. Holding social factors constant, they exposed listeners to a subtle accent created by altering the distributional regularities of acoustic speech dimensions relative to American English norms (Idemaru & Holt, 2011). Replicating prior research, exposure to the accent changed the acoustic dimensions listeners relied upon in speech perception (Idemaru & Holt, 2011; Hodson et al., 2023). Additionally, the acoustic dimension perceptually down-weighted in the context of the accent was also less distinctive in listeners' own speech productions. In short, statistical learning across distributional patterns of an accent shifts both perception and production.

The scope of processes that shift perceptual representations in a manner that influences production is thus quite broad: statistical learning, explicit training with feedback, and contextual interactions each gives rise to changes in speech production. Yet, contemporary models of speech motor control have been built to model adjustments to articulation driven by sensory feedback from one's own voice (e.g., Guenther, 2016) and do not directly account for these influences of external speech. In these neurobiologically plausible accounts, internal feedforward models predict the sensory outcomes of articulation and continuously evaluate incoming sensory feedback against the predictions. Mismatches create error that updates speech motor representations in a feedforward manner, correcting articulation (Lametti, Nasir, & Ostry, 2012; Larson, Altman, Liu & Hain, 2008). It would be parsimonious for the influences of perception of others' speech on speech motor control to interface with these same neurobiological systems.

However, although the perceptual representations guiding speech motor control through sensory feedback appear to be more malleable than previously thought (Houde & Jordan, 1998; Purcell & Munhall, 2006; Rochet-Capellan & Ostry, 2011; Tourville et al., 2008; Villacorta et al., 2007), they are not well-specified in current models. Here, we ask: what is the nature of the representations that underlie perceptual effects of other voices on speech production? Prominent theoretical accounts of phonetic convergence posit alignment of talkers through detailed phonetic encoding in articulatory terms (e.g., Pickering & Garrod, 2013; Shockley et al., 2004), a tenet related to theories of speech perception dependent on the parsing of articulatory-phonetic features (Fowler, 2006; Liberman et al., 1967). This would predict that articulatory-phonetic information should be vital in causing perception of others' speech to influence speech production.

We examine this prediction by capitalizing on the powerful influence of preceding context on speech *perception* to study its influence on speech *production*. Previous perception research has shown that preceding syllables (Holt et al., 2000; Lindblom & Studdert-Kennedy, 1967; Lotto & Kluender, 1998; Mann, 1986), sentences (Huang & Holt, 2012;

Ladefoged & Broadbent, 1957; Laing et al., 2012; Stilp & Assgari, 2018) and even nonspeech tones (Holt, 2005, 2006a, 2006b; Lotto & Kluender, 1998) shift perception of subsequent speech. These effects are reliably contrastive, with the spectrotemporal characteristics of a precursor sound pushing perception of subsequent speech in an opposing direction. For example, high-frequency context sounds tend to cause subsequent speech targets to be perceived as possessing lower-frequency acoustic energy, as in the perceptual shift from /ɛ/ to /I/ caused by higher-frequency carrier phrases in the Bourguignon et al. (2016) study described above.

Spectrally contrastive context effects present an opportunity to scrutinize the perceptual representations that underlie sensorimotor effects in speech because they arise across distinct listening contexts that vary in the extent to which they convey articulatory-phonetic information. Across three studies, we examine the acoustics of participants' /ga/-/da/ speech productions in distinct listening contexts. In Experiment 1, syllabic speech contexts differ in detailed articulatory-phonetic information. In Experiment 2, continuous speech audiobook contexts vary in overall their spectral profile, but articulatory-phonetic information is constant. In Experiment 3, nonspeech contexts composed of sequences of tones lack articulatory-phonetic information. Though these contexts differ substantially, each is defined by a manipulation of the underlying spectral energy in a manner that we expect to produce spectrally contrastive context effects on /ga/-/da/ perception based on prior studies. We ask which, if any, of these perceptual shifts influences speech production.

In Experiment 1, we pair preceding /al/ and /ar/ context syllables with perceptually ambiguous /ga/-/da/ target syllables. Preceded by /al/, perceptually ambiguous /ga/-/da/ target syllables tend to be more often reported as /ga/ whereas the same targets are more often categorized as /da/ following /ar/ (Lotto & Kluender, 1998; Mann, 1980, 1986). This perceptual context effect, and others like it (e.g., Lindblom & Studdert-Kennedy, 1967; Mann & Repp, 1980; 1981), have played an important role in theories of speech perception because they appear to compensate for coarticulation, the acoustically assimilatory influence of articulation across adjacent speech sounds. This might involve parsing articulatory-phonetic features (Fowler, 2006; Liberman et al., 1967). Or, alternatively, the perceptual effects might be driven by general auditory processes that exaggerate spectral contrast: /al/ (with *higher*-frequency F3) shifts target categorization toward "ga" (with *lower*-frequency F3) and /ar/ (with *lower*-frequency F3) leads the same ambiguous target to more often be reported as "da" (with *higher* F3 energy) (Lotto & Kluender, 1998; Mann, 1986). Consistent with a general perceptual account, even Japanese quail (*Coturnix coturnix japonica*) trained to peck keys in response to /ga/ and /da/ exhibit perceptual context effects in the same direction as human listeners (Lotto et al., 1997). Experiment 1 tests whether the influence of /al/ and /ar/ contexts on /ga/-/da/ perception also affects speech production.

Experiment 2 tests whether naturalistic speech contexts differing in acoustic spectrotemporal profile, but not in articulatory-phonetic detail, produce perceptual context effects that transfer to speech production. Previous research has shown that perceptual context effects can arise from acoustic shifts such as those arising from variation in room reverberation (Watkins, 1991; Watkins & Makin, 2007). Experiment 2 asks whether these perceptual context effects produce changes to speech production. Participants listen to excerpts from

Harry Potter and the Sorcerer's Stone (Rowling, 1998) filtered to create two versions of the audiobook that vary subtly in spectral energy in the regions that correspond roughly to the spectral differences between /al/ and /ar/ in Experiment 1. If resolving another talker's speech in detailed articulatory-phonetic terms is necessary to support transfer to speech production, we expect Experiment 2 to yield perceptual context effects driven by spectral contrast that do not transfer to speech production. However, if general auditory processes underlying spectral contrast effects on speech perception are sufficient to support transfer to production, we predict acoustic differences in /ga-/da/ utterances as a function of listening context.

Experiment 3 provides an even stronger test of the necessity of articulatory-phonetic information in transfer of perceptual context effects to speech production. Prior research establishes a spectrally contrastive influence of sequences of nonspeech tones on speech perception (Holt, 2005, 2006b, 2006a). If articulatory-phonetic information is essential for inducing shifts in speech production, then Experiment 3 should replicate this perceptual context effect, but we should observe no influence on speech production.

To summarize, if general auditory interactions are sufficient to influence speech motor control, then we should observe effects of context on both *perception and production* in each experiment. Alternatively, if detailed articulatory-phonetic encoding of another talker's speech in articulatory terms is necessary to support transfer to production (Pickering & Garrod, 2013; Shockley et al., 2004), then we predict perceptual context effects, but not transfer to production, in Experiment 3 where the simple nonspeech context lacks articulatory cues. Experiment 2 presents an important intermediate test case because continuous speech contexts possess articulatory-phonetic information, but this information does not vary across conditions. Instead, the Experiment 2 conditions vary non-specifically in spectrotemporal acoustic profile, akin to how room acoustics subtly influence the overall quality of a recording. If detailed phonetic encoding is necessary for transfer to speech production, then Experiment 2 should reveal a speech-driven perceptual context effect that does not transfer to production. In this way, we leverage perceptual context effects to better understand the nature of the perceptual representations that interact with speech production.

Experiment 1

Experiment 1 asks whether the influence of /al/ and /ar/ contexts on /ga/-/da/ perception also affects speech production.

Methods

Transparency and Openness: Stimuli, data and code are publicly available on Open Science Framework at https://osf.io/mnj85. The studies were not preregistered. As a first study to examine production in these listening contexts, we took a conservative approach that involved: (1) within-participants designs that minimize across-talker speech variability; (2) consistent target speech syllables to elicit categorization responses and productions across conditions and experiments; (3) a feature-agnostic acoustic measure of production; (4) conservative statistical analyses conducted across the entire frequency spectrum (instead

of point-wise) and validated with control analyses on samples drawn without respect to condition.

Participants: We examined perception and production across 55 adult native-English participants residing in the United States and reporting normal hearing (20–35 years; M = 28.3, SD = 4.2; 29 Female, 22 Male, 3 Non-binary, 1 Prefer Not To Answer, with responses chosen from a drop-down list). All participants were recruited for online testing using Prolific (www.prolific.com). Without prior studies to inform effect size, we targeted this sample to yield approximately 13,000 productions (~3,000/condition) to enter our test of transfer to speech production.

To arrive at this sample, we excluded data from 27 additional online participants. Six failed to follow task instructions and one had an issue with audio recording. An additional 20 participants responded exclusively /ga/ or /da/ in at least one condition, precluding analysis of across-category acoustic differences in speech production. We anticipated exclusion on this basis because we selected the two target syllables based on aggregate perceptual effects from prior studies. Thus, we expected that individual differences in phonetic category boundaries would lead some participants to report only /ga/ or /da/. Even with these exclusions, target selection based on aggregate perceptual boundaries is the preferrable approach because by-participant target selection would confound target syllable acoustics with our manipulation of listening context.

Stimuli: Each trial involved one of two context syllables, /ar/ or /al/, and one of two target syllables chosen to be acoustically ambiguous between /ga/ and /da/. The /al/ and /ar/ context syllables were synthesized using the cascade branch of the Klatt speech synthesizer (Klatt, 1980) as described by Stephens and Holt (2003). In brief, a 100-ms steady-state /a/ vowel was followed by a 150-ms linear formant transition associated with the final consonant, for 250-ms total syllable duration with the third format (F3) offset frequency differentiating /al/ and /ar/. For /ar/, F3 frequency decreased from the steady-state vowel (2440 Hz) to syllable offset (1593 Hz) whereas for /al/ F3 frequency increased from 2440 Hz to 2863 Hz. Formant frequencies modelled a monolingual American English male talker (Lotto & Kluender, 1998).

The two target syllables originated from the same talker's /ga/ and /da/ productions. With natural recordings as endpoints, stepwise morphing of onset frequencies created approximately equal steps across nine 589-ms syllables (Analysis-Synthesis Laboratory, Kay Elemetrics, Lincoln Park, NJ), as described in prior studies (Holt, 2005, 2006b; Laing et al., 2012; Wade & Holt, 2005). Based on /ga/-/da/ categorization responses across the full 9-step series, we chose two ambiguous target syllables: one more acoustically /ga/-like and one more /da/-like (syllables 4 and 6, respectively, along the 9-step series in Figure 2 of Holt, 2005). Context and target syllables possessed RMS-matched amplitude.

Procedure: We capitalized on the spectrally contrastive influence of context effect to selectively pair context syllables (/al/, /ar/) and perceptually ambiguous target syllables (/ga/-like, /da/-like) as in Stephens and Holt (2003). Based on prior studies, pairing /al/ with the more /ga/-like target is expected to shift perception of this ambiguous target toward

"ga" and pairing /ar/ with the more /da/-like target should shift its perception toward "da" (Lotto & Kluender, 1998; Mann, 1980, 1986). Thus, the net influence of this paring is to *enhance* target syllables' perceptual distinctiveness. In contrast, the opposite pairing should *diminish* target syllable distinctiveness, as illustrated in Figure 1A. In this way, leveraging the perceptually contrastive directional influence of preceding contexts allowed us to selectively pair identical context and target syllables in a manner expected to produce a perceptual shift in target-syllable categorization. Stimuli were identical across conditions; only trial-wise syllable pairings differed.

The Gorilla software platform (gorilla.sc; Anwyl-Irvine et al., 2020) hosted the online experiment, with the restriction that participants use a desktop or laptop computer (no phones, tablets), the Google Chrome browser, wired headphones, and a microphone. Prior to beginning the experiment, participants set the system volume to a comfortable level for a continuous white noise, with instructions not to adjust this level during the experiment. Next, a simple task assured headphone use (Milne et al., 2021). Finally, a microphone check ensured successful audio recording. Participants failing either system check did not enter the study.

As shown in Figure 1B, each trial involved one of the two context syllables (/ar/ or /al/) followed by a 50-ms silent interval and one of the two target syllables. After 300 ms, a microphone icon appeared on the screen, prompting participants to utter the target syllable. Recording continued for 2 sec and utterances were stored digitally for subsequent acoustic analysis. Immediately after, on-screen response alternatives 'G' and 'D' prompted participants to report the syllable they had produced with a keypress. The next trial commenced 750 ms later. Before starting the experiment, participants performed three practice trials and received written instructions to repeat the target syllable upon seeing the microphone icon and to respond with a keypress indicating the syllable they heard upon seeing the on-screen response alternatives. Each participant experienced both Enhanced (120 trials) and Diminished (120 trials) conditions with self-paced breaks after every 40 trials, and condition order counterbalanced across participants.

Analytic Approach

Perceptual Categorization. We tested perceptual categorization responses using a generalized linear mixed effects model implemented using the *glmer* function of the *lme4* package in R (Bates et al., 2015 in R, version 4.1.3; R Core Development Team, 2022). We used a mixed-effects logistic regression model with a binary perceptual categorization response of /ga/ or /da/ as the dependent variable. The full statistical model included fixed effects across Condition (Enhanced, Diminished), Target Syllable (/ga/-like, /da/-like), and their interaction. We included a random intercept for subject and describe the most complete model tolerated; no model converged with random slopes. Analyses excluded a minority of trials for which the perceptual categorization response occurred more than 3 sec after target syllable offset (1.1% of total trials) and trials for which there were audio recording issues that precluded acoustic analysis (2.2% of total trials); we excluded both perception and production data for these trials.

Acoustic Analysis of Productions. Prior research guides specific, directional predictions about the influence of the listening contexts (conditions) on perception. As a first test of whether these perceptual influences transfer to production, we adopted a feature-agnostic, global measure of acoustic differences across conditions.

We accomplished this by examining the absolute value of the difference in spectral energy for /ga/ versus /da/ utterances. This difference across the long-term average spectrum (LTAS), a measure of overall acoustic energy across frequency, provides a sparse representation of speech acoustics that allows us to test whether listening contexts (and their perceptual consequences) produce global shifts in spectral energy in speech production.

We used the 'to TextGrid (silences)...' function in Praat (Boersma & Weenink, 2025) to isolate each utterance. We next used a voice onset time (VOT) detection algorithm (Dr. VOT; Shrem et al., 2019) to extract the VOT. This provided a common reference point (the onset of voicing for productions with a positive VOT or the onset of the plosive burst for tokens with a negative VOT) to mark syllable onset. From onset we trimmed productions to 75 ms, a temporal window closely associated with the initial consonant. We next amplitude-normalized these sounds and calculated the LTAS of each (Praat 'to LTAS…' function with 100 Hz bandwidth, focusing on frequencies to 5,000 Hz). This analysis pipeline required an utterance of at least 150 ms; we excluded shorter utterances from analyses.

This approach confers several advantages. Above all, it sidesteps an issue that has been challenging to studies of phonetic convergence (see Ostrand & Chodroff, 2021): which of the many acoustic-phonetic dimensions that might be affected should be monitored for change? Adopting a feature-agnostic approach across the full acoustic frequency spectrum eliminates guesswork regarding the putative features that may be affected by listening context and allows us to utilize the same acoustic measure across experiments.

This may be especially important for assessing listening-context-dependent changes in the acoustic signatures of our /ga/-/da/ target syllables, for which the onset frequencies of the second (F2) and third (F3) formants typically vary. Taking these putative acoustic landmarks as examples, the articulation of /ga/ results in F2 and F3 onset frequencies situated close in frequency, with energy prominences overlapping in frequency with a limited 'trough' separating them. Articulation of /da/, in contrast, results in more distinct F2 and F3 prominences. Consideration of how these representative acoustic landmarks might be influenced by contexts illustrates two general challenges for assessing the influence of listening context on production: (1) talker variability and (2) directionality of influence.

Talkers introduce variability in part due to vocal tract length, with shorter vocal tracts producing higher formant frequencies (Stevens, 1998). Thus, a talker with a shorter vocal tract may produce an F2 frequency in approximately the same spectral region as the F3 of a talker with a longer vocal tract. In this way, idiosyncratic talker differences introduce ambiguity in group analyses. With acoustic data mixed across talkers, an observed change in spectral energy might be attributable to a change in F2, a change in F3, or collaborative influence pushing the putative acoustic features in opposing directions. The same issue complicates directional acoustic predictions. If listening context were to push one region of

the acoustic spectrum in one direction and another region in an opposing direction there may be complex additive and/or subtractive effects that are further complicated by talker variability.

For these reasons, we examine the full spectrum of acoustic energy across the first 75 ms of each utterance, before vowel onset and where acoustic differences between /ga/ and /da/ can be expected to be reliably concentrated. By measuring the long-term average spectrum (LTAS) across trials for which participants labelled their response as /ga/ versus /da/, we examine spectrum-wide changes. This circumvents assignment of targeted regions of spectral energy to putative features. We report the absolute value of the LTAS difference between utterances from ga-response trials and utterances from da-response trials. This allows us to quantify overall changes in acoustic energy that are resilient to push-pull interactions that can be expected to be present in acoustic data pooled across talkers. We note that while this approach addresses some expected challenges, measurement of the absolute value of acoustic differences precludes directional predictions. This sacrifice is acceptable because the primary aim of the present experiments is to identify which, if any, of the diverse listening contexts spanning syllables, audiobooks, and tones that produce perceptual context effects also influence production.

Complementing this, we undertake a control acoustic analysis for which we apply the same acoustic analysis pipeline to ga-response trials and da-response trials sampled without regard to the listening context in which the speech was uttered. This establishes a baseline expectation for the statistical likelihood of speech variation not attributable to listening context to be the driver of any production effects we observe across listening contexts.

Statistical Analysis of Productions. We conditioned production analyses on the perceptual categorization response from the same trial from which an utterance was recorded. Since perceptual categorization varied according to listening context by design, there were uneven numbers of /ga/ versus /da/ utterances across conditions. We thus used nonparametric sampling and cluster-based permutation analysis to accommodate differences in observations across conditions and to account for multiple comparisons, respectively. We randomly sampled 100 /ga/-response trials and 100 /da/-response trials across all participants, drawn from the same (Enhanced, Diminished) condition and computed the LTAS Difference, defined as the absolute value of the difference in spectral energy of utterances categorized as /ga/ versus /da/ across 0–5,000 Hz. We repeated this process 100 times/condition.

As a control analysis, we applied the same analysis pipeline to a random sample of 100 /ga/-response trials and 100 /da/-response trials, made without regard to listening context. We then repeated this procedure a second time. Statistical comparison of LTAS Differences across /ga/-response and /da/-response trials for these two samples provided a test of whether any production differences observed across Enhanced and Diminished contexts could arise from acoustic variability typical across speech productions, without an influence of the listening context *per se*.

For both control samples and those sampled according to condition, we estimated confidence bands that included 95% of *all LTAS curves across all frequencies* using functional data analysis (Degras, 2017). Note that this differs from calculating frequency-wise confidence bands pointwise along the frequency spectrum. In virtually all situations, functional data analysis across the full curve is the more conservative estimate (Degras, 2017).

We then tested the statistical significance of the LTAS Difference across conditions using cluster-based permutation analysis (Maris & Oostenveld, 2007) implemented by the *clusterlm* function in the *permuco* R package (Frossard & Renaud, 2021). We report cluster mass as the sum of t over all frequency bands in the cluster with p values reported as the probability that a cluster appears randomly in the permutation test (i.e., the proportion of clusters with a larger cluster mass uncovered through random permutations of the data). This test reveals whether differences exist between the two listening contexts (or separate samples in the case of the control analysis) across the frequency spectrum and whether the size of clusters is significant. However, since the statistical test is performed over cluster sizes irrespective of location along the curve the approach does not establish precise onsets and offsets of the frequency ranges that differ between the two conditions (Sassenhagen & Draschkow, 2019). In the primary analysis across conditions, all significant clusters indicate an influence of the perceptual context effect on production because context and target sounds are identical across conditions; only the selective pairing of contexts with targets varies.

Results

Perceptual Categorization: We first examined whether the context syllables influenced target syllable categorization, replicating 'compensation for coarticulation' effects (Lotto & Kluender, 1998; Mann, 1980; Repp & Mann, 1981). Figure 2A plots the proportion "ga" responses as a function of Target Syllable and Condition. There is a significant main effect of Target Syllable in the expected direction: the perceptually ambiguous /ga/-like syllable was more often categorized as /ga/ than the /da/-like syllable ($z = -20.45$, $p < .001$). There was also a main effect of Condition. Participants categorized target syllables more often as /ga/ in the Diminished, compared to the Enhanced, condition ($z = 6.18$, $p < .001$). Most important, the strategic pairing of context and target syllables to Enhance or Diminish perceptual distinctiveness influenced perceptual categorization, as revealed by a significant interaction ($z = -11.08$, $p < .001$). The pattern of target syllable categorization differed significantly as a function of preceding context. Specifically, the directionality of influence replicates prior 'perceptual compensation for coarticulation' studies (Holt & Lotto, 2002; Lotto & Kluender, 1998) such that categorization of target syllables was less distinct in the Diminished condition. Table I shows the full statistical analysis.

Speech Production: We next investigated whether this perceptual context effect influenced speech production. As evident in Figure 2B, cluster-based permutation tests revealed a significant influence of Condition (Maximum Cluster Mass, $t_{max}(198) = 466.05$, $p < .001$) apparent across five clusters (see Table II). Two of these clusters lie within the region of the spectrum most important in distinguishing /ga/ and /da/. In the region approximately

bounded by 1,500–1,900 Hz there was a significant difference in spectral power between the Enhanced and Diminished condition (Cluster Mass, $t(198) = 55.99$, $p = .016$), as well as in the region approximately bounded by 2100–3700 Hz (Cluster Mass, $t(198) = 466.05$, $p < .001$). Additionally, there were significant differences at lower frequencies (400 Hz; Cluster Mass, t(198) = 50.38, p = .020; 700–800 Hz; Cluster Mass, $t(198) = 39.57$, $p = .034$) and higher (4,500–5,000 Hz; Cluster Mass, $t(198) = 87.23$, $p = .005$) frequencies.

These results are best evaluated against the control cluster-based permutation test for which we calculated LTAS differences over 100 /ga/ and 100 /da/ utterances sampled without regard to condition. As shown in Figure 2C, there were no statistically significant differences.

Discussion

To summarize, the /al/ and /ar/ syllables of Experiment 1 established propitious conditions for perceptual context effects to transfer to speech production because they carry strong articulatory-phonetic information. In fact, this is what we observed: perceptual context effects driven by another talker's speech influenced participants' own speech productions.

Experiment 2

Capitalizing on the Experiment 1 approach, we next asked whether continuous speech contexts differing in acoustic, but not articulatory-phonetic, details produce perceptual context effects that transfer to speech production.

Methods

Participants: We tested 54 adult native-English participants residing in the United States (18–35 years, $M = 26.9$, $SD = 4.9$; 22 Female, 30 Male, 2 Non-binary, with responses chosen from a drop-down list) with normal hearing recruited using Prolific (prolific.co). We excluded data from an additional 16 participants; three did not perform the task according to instructions, four had audio recording difficulties, and nine responded only /ga/ or /da/ in at least one condition.

Stimuli: The Experiment 2 design followed Experiment 1, with Enhanced and Diminished conditions intended to shift categorization of the two Experiment 1 target syllables. In Experiment 2, the context stimuli were short excerpts extracted from a 6 minute 33 second recording from the first chapter of *Harry Potter and the Sorcerer's Stone* (Rowling, 1998) spoken by the same adult male American-English talker after whom the target syllables were modelled (from "*Mr. and Mrs. Dursley, of number four, Privet Drive, were proud to say that they were perfectly normal, thank you very much*" to "*When Dudley had been put to bed, he went into the living room in time to catch the last report on the evening news*"). We extracted 120 excerpts ranging from 1.7 to 5.0 sec ($M = 3.0$ sec, $SD = 0.7$) from the recording, with offsets positioned at clause and sentence breaks so as not to unduly disrupt the flow of the story.

We applied two filters (Adobe Audition) over the 6 minute 33 second audiobook, resulting in two audiobooks with distinctive long-term average spectra that cross in power at

approximately 2,300 Hz. The net effect of this was to emphasize spectral energy in lower (Low LTAS) or higher (High LTAS) frequency regions in the range of approximately 1,800–2,800 Hz. This subtle acoustic manipulation did not affect perceived talker identity, accent or speaking style, and did not impact speech intelligibility (see also Bosker et al., 2020). The net result of filtering was subtle, as if passages were recorded in different rooms. Most importantly, filtering did not affect the articulatory-phonetic information available to listeners across the continuous natural-speech passages. There were no systematic shifts of the word, phoneme, or phonetic feature acoustics (e.g., as in Maye et al., 2008). The audiobook passages were matched in RMS amplitude.

Procedure: As in Experiment 1, we paired the Low and High LTAS audiobooks with the two perceptually ambiguous /ga/-/da/ target syllables to Enhance or Diminish targets' perceptual distinctiveness according to predictions from spectral contrast. In the Enhanced condition, we paired the High LTAS audiobook with the more /ga/-like target syllable and the Low LTAS audiobook with the more /da/-like target syllable (Figure 2A). We expected the opposite pairing in the Diminished condition to shift perception in the opposing direction, reducing perceptual distinctiveness of the same target syllables. As in Experiment 1, identical context and target syllables defined the Enhanced and Diminished conditions; only trial-wise parings differed.

On each trial, participants listened passively to a short (1.7 to 5.0 sec) audiobook excerpt. After a 50-ms silent interval, one of the two target syllables followed. A 2-sec microphone icon on the screen prompted participants to repeat the target and then 'G' and 'D' labels prompted participants to categorize the target (Figure 2B). The audiobook excerpts progressed sequentially so participants could follow the story. Half of the excerpts (60 trials) were followed by the /ga/-like target syllable and half by the /da/-like target syllable, with order randomized. The order of the 120-trial Enhanced and Diminished conditions was counterbalanced across participants such that each participant heard the audiobook twice. There were self-paced breaks every 40 trials. In other ways, Experiment 2 followed the approach of Experiment 1.

Analytic Approach: We followed the Experiment 1 analysis pipeline, excluding a minority of trials due to recording issues (0.9% of total trials) or delayed (> 3 sec) perceptual response (2.5% of total trials). We excluded both perception and production data from these trials.

Results

Perceptual Categorization: The subtle shift in spectral energy experienced across continuous speech influenced target syllable categorization. As shown in Figure 3A, participants categorized the /ga/-like target syllable significantly more often as /ga/ than the /da/-like target ($z = 3.03$, $p = .002$) and, overall, there were more /ga/ responses in the Enhanced condition ($z = 29.27$, $p < .001$). Most important to our predictions, these factors interacted ($z = -37.35$, $p < .001$). The nature of this interaction was consistent with spectral contrast accounts (Holt & Lotto, 2006; Huang & Holt, 2012; Laing et al., 2012; Lotto & Kluender, 1998) that shaped design of the Enhanced and Diminished conditions. The

same target syllables were categorized more distinctively in the Enhanced, compared to the Diminished, condition. Table III provides the full statistical analysis.

Speech Production: We next asked whether this perceptual context effect transfers to speech production. As is evident in Figure 3B, cluster-based permutation tests revealed a significance difference in speech productions across the Enhanced and Diminished listening contexts (Maximum Cluster Mass, $t_{max}(198) = 107.12$, $p = .002$), apparent in two clusters spanning approximately 800–1200 Hz (Cluster Mass $t(198) = 107.12$, $p = .002$) and 2,600–3,900 Hz (Cluster Mass $t(198) = 86.99$, $p = .005$). Table IV shows the full cluster analysis. As shown in Figure 3C, a control analysis with speech productions sampled without regard to condition revealed no significantly different clusters. This reassures that the differences in speech production shown in Figure 3B arose as a function of listening context.

Discussion

In summary, continuous audiobooks filtered to possess different profiles of spectral energy as if recorded in different rooms shift speech perception in a manner consistent with spectral contrast. This filtering did not affect the articulatory-phonetic information available to listeners across the continuous natural-speech passages. There were no systematic shifts of the word, phoneme, or phonetic feature acoustics. Yet, the perceptual effect of context carried over to influence production.

Experiment 3

Experiment 3 examined whether listening to a sequence of tones, entirely lacking articulatory-phonetic information but situated to convey acoustic energy in the spectral region differentiating /al/ and /ar/ in Experiment 1 would produce perceptual context effects that affect speech production.

Methods

Participants: We tested 57 adult native-English participants (18–35 years, $M = 28.3$, $SD = 4.3$; 28 Female, 29 Male, with responses chosen from a drop-down list) with normal hearing and residing in the United States and recruited using Prolific (prolific.co). We excluded three additional participants for responding only /ga/ or /da/ in at least one condition.

Stimuli: The two target syllables were identical to Experiments 1 and 2. Precursor contexts modelled the approach of Holt (2005), with sequences of 21 70-ms sinewave tones (30-ms ISI, 5-ms linear onset/offset ramp) each possessing a unique frequency drawn from a spectral distribution defined ±500 Hz around a mean frequency (50-Hz steps). All 21 possible tones in the frequency range were played once per trial, with tone order randomized across trials. A High LTAS condition was defined by tone sequences sampled around a mean frequency appropriating the higher F3 frequency of /al/ (2,800 Hz). A Low LTAS condition was defined by a mean approximating the F3 frequency of /ar/ (1,800 Hz). Stimuli were created using MATLAB (The MathWorks Inc., 2024).

Procedure: Following the approach of Experiments 1 and 2, High LTAS tone sequences preceded the /ga/-like target syllable and Low LTAS sequences preceded the /da/-like target syllable in the Enhanced condition (Figure 2A). The opposite pairing defined the Diminished condition. Each condition involved 120 trials, with breaks every 40 trials. Condition order was counterbalanced across participants.

In contrast to Experiments 1 and 2, participants first categorized the syllable and then uttered it (Figure 2B). In Experiment 3 participants passively listened to a 2.1 sec tone history and, after a 50-ms silent interval, heard one of the two target syllables. A visual prompt signalled participants to categorize the target syllable with a key press corresponding to “G” or “D”. After a 300-ms pause, the same target syllable repeated, and a 2-sec microphone icon prompted participants to utter the syllable.

There are three reasons for this subtle change. First, we sought to examine the robustness of the Experiment 1 and 2 effects to task demands. Second, we reasoned that moving production to the end of the trial would provide a more conservative test of transfer to production due to the greater temporal delay of production from the condition-defining nonspeech precursors. Third, this temporal delay along with the repeat of the target syllable minimizes transfer from being wholly attributable to lingering perceptual activation or mid-level perceptual grouping effects across speech and nonspeech sound sources.

Analytical Approach: We followed the approach of Experiment 1, excluding a minority of trials due to recording issues (0.2% of total trials) or slow perceptual responses (>3 sec, 3.5% of total trials). Excluded trials were omitted from both perceptual categorization and speech production analyses.

Results

Perceptual Categorization: We first asked whether the tone sequences affected speech categorization. As shown in Figure 4A, our results replicated prior studies (Holt, 2005, 2006a, 2006b). We observed a main effect of Target Syllable, with greater /ga/ categorization responses for the acoustically more /ga/-like ambiguous token ($z = -6.13$, $p < .001$). As in Experiments 1 and 2, there was a significant tendency to respond /ga/ more often in the Enhanced condition ($z = 11.26$, $p < .001$). Of most interest, Target Syllable and Condition significantly interacted ($z = -16.8803$, $p < .001$). The spectral energy present across a sequence of sinewave tones carrying no speech information shifted categorization of subsequent /ga/-/da/ syllables. The full statistical analysis is shown in Table V.

Speech Production: Do perceptual context effects driven by nonspeech tones transfer to speech production? As shown in Figure 4B, we observed significant differences in speech production across conditions (Maximum Cluster Mass, $t_{max}(198) = 390.47$, $p < .001$). These differences emerged across four clusters. Two of these clusters lie within the region of the spectrum important in distinguishing /ga/ and /da/. In the region approximately bounded by 2,100–2,300 Hz there was a significantly greater spectral energy difference in the Enhanced, compared to the Diminished condition (Cluster Mass $t(198) = 37.46$, $p = .033$) whereas the spectral energy difference was greater in the Diminished condition in the 2,600–4,000 Hz region (Cluster Mass $t(198) = 390.47$, $p < .001$). Additionally, a lower-frequency cluster

(1,200–1,500 Hz; Cluster Mass $t(198) = 63.19$, $p = .009$) was also significantly different across conditions. Table VI shows the full cluster analysis. As shown in Figure 4C, a control analysis with speech productions sampled without regard to condition revealed no significantly different clusters across spectral power.

Discussion

Experiment 3 provides the strongest test for the involvement of general auditory processes in perception-production interactions in speech. As in Experiments 1 and 2, we observe that spectrally contrastive perceptual context effects driven by exposure to sequences of tones carry over to speech production.

General Discussion

What kinds of representations underlie the perceptual effects of other voices on speech production? To examine this question, we investigated whether shifts in speech perception triggered by various preceding acoustic contexts carry over to influence speech production. Experiment 1 demonstrates that the perceptual influence of preceding syllables conveying strong articulatory-phonetic information affects production. Experiment 2 shows that broader shifts in the spectral profile of continuous speech that do not impact articulatory-phonetic information also affect both perception and production. Most strikingly, Experiment 3 demonstrates that sequences of nonspeech sinewave tones lacking any linguistic content induce perceptual shifts that influence speech production. Together, these findings indicate that seemingly ‘low level’ general auditory perceptual processes that sharpen spectral contrast across successive sounds exert an influence on speech motor control.

While spectral effects on speech *perception* have been well established, to our knowledge, this is the first demonstration that they affect speech *production.* Across three experiments, we strategically paired acoustic contexts with target syllables to create a paradigm designed to either enhance or diminish speech targets’ perceptual distinctiveness. Prior research has consistently demonstrated that spectral properties of context sounds can bias speech *perception*: contexts with higher-frequency energy tend to shift perception toward lower-frequency target syllables, and vice versa (e.g., Holt, 2005, 2006b; Holt & Lotto, 2006; Laing et al., 2012; Lotto & Kluender, 1998; Stilp & Assgari, 2018). Notably, this effect arises from general auditory processes rather than speech-specific mechanisms and can be triggered by nonspeech sounds that mimic the spectral characteristics of speech, as replicated in Experiment 3 (Holt, 2005, 2006a, 2006b; Huang & Holt, 2012; Laing et al., 2012). It has even been observed in nonhuman animals (Bartlett & Wang, 2005; Lotto et al., 1997).

These effects are robust. They persist even when context and target are presented to opposite ears (Lotto et al., 2003) and across temporal gaps exceeding one second (Holt, 2005), pointing to central auditory, likely cortical, mechanisms. Holt (2006b) makes a case that spectrally contrastive perceptual shifts in speech perception may arise from stimulus-specific adaptation in the auditory system prompting shifts in the neural population that encodes subsequent speech (Ulanovsky et al., 2004; see Song et al., 2023 for a review). The novel

insight from the present work is that these general auditory interactions extend beyond perception to actively shape speech production, as well.

This offers both methodological and theoretical insights. On the methodological side, our technique has several advantages. Pairing targets strategically with listening contexts according to directional predictions from spectral contrast effects allowed us to hold the acoustics of both context sounds and target syllables constant across conditions. This isolated perceptual consequences of these pairings as the drivers of effects on production. Complementing this, we adopted a feature-agnostic approach to acoustic analyses. Examining global energy across the full acoustic spectrum afforded us the chance to uncover effects that might be missed in targeted examination of specific acoustic landmarks or features. As described above, it is important to note that while this measure allows us to observe systematic differences in speech productions as a function of listening context, it does not support directional predictions. We reasoned that cross-talker differences like vocal tract length could play off one another to masquerade as directional differences that are, in fact, serendipitous additive/subtractive effects of idiosyncratic energy prominences in speech. Our reliance on the absolute value of the difference in long term spectral energy of large samples of /ga/ versus /da/ utterances circumvents this issue but does limit conclusions about the precise nature of articulatory differences across conditions, or whether the productions were potentially more ‘distinctive.’

Even so, by demonstrating that spectral contrast effects transfer to production, the present studies provide a foundation for future work to test directional predictions. Nonspeech listening contexts, across which it is possible to parametrically manipulate spectral energy with fine precision, are likely to be especially revealing. In this regard, it is notable that condition-wise differences in the long-term average spectrum of /ga/ versus /da/ reveal sensible patterns. Significant clusters consistently emerged in the spectral region corresponding to second and third formant frequencies, key acoustic cues for /g/-/d/ categorization (Delattre et al., 1955).

Yet, there were differences across experiments. Interestingly, the punctate syllable (Experiment 1) and nonspeech tone (Experiment 3) contexts influenced production in a highly similar manner. For each, the absolute value of spectral energy differences between /ga/ and /da/ was greater in a lower frequency region for utterances in the Enhanced condition and greater in a higher frequency region for utterances in the Diminished condition. These “low” and “high” regions of the spectrum are broadly consistent with second (F2) and third (F3) formant frequencies for /ga/ and /da/, with the caveats regarding talker differences we described above. Most important, the similarity of acoustic patterns across Experiments 1 and 3 is instructive to theory: syllables carrying strong articulatory-phonetic information and tones carrying none affect speech production in a similar manner.

As described above, the complexity, talker-variability, and combinatorial nature putative acoustic landmarks makes simple directional predictions across LTAS difficult. However, this does not mean that the observed directions are arbitrary. The ‘flip’ in directionality across the two spectral regions in Experiments 1 and 3 is consistent with the push-pull effects that we designed our analyses to protect against and thus emphasize the importance

of our feature- and directionality-agnostic approach to acoustic analyses. Experiment 1 and 3 contexts possess concentrated, local spectral energy. In contrast, the natural speech contexts from the audiobook of Experiment 2 involved a spectrum-wide acoustic manipulation. In this context, we can speculate that the opposing push-pull effects we observe in Experiments 1 and 3 may be interacting with the more complex, spectrum-wide acoustic differences of the natural speech contexts of Experiment 2. We speculate that the broader spectral profile differences across Experiment 2 continuous speech contexts led to more complex perceptual interactions than measured in our two-alternative perceptual categorization task. This possibility has some support in prior research. Spectral contrast effects on speech can be modulated with fine-grained manipulation of spectral energy across context sounds (Holt, 2006b). The similarity of influence of syllables and tones on speech motor control establishes that future studies can leverage psychophysical manipulation of nonspeech tone distributions carrying increasingly complex patterns of spectral energy to better understand how speech motor control is affected by perceptual shifts.

Finally, we note that we conducted our analysis as a function of the intended utterance, as gleaned from /ga/-/da/ trial-wise perceptual responses. This choice eliminated the potential for the reduced perceptual distinctiveness expected in the Diminished condition to be mistaken for an effect of listening context. Our control analyses underscore that the differences we report do indeed arise from listening context; there were no significant differences in speech production across /ga/ and /da/ utterances sampled without regard to context. This assures us that the influence of listening context is not simply a result of natural acoustic variability across speech productions.

On the theoretical side, our findings offer new insight into the mechanisms linking speech perception and production. Notably, the observed effects cannot be attributed to social factors typically associated with phonetic convergence (Bourhis & Giles, 1977; Gregory & Webster, 1996; Michalsky & Schoormann, 2017), as each experiment used a single voice (or tones) and non-interactive tasks. Covert mimicry, a proposed driver of phonetic alignment between talkers (Pickering & Garrod, 2013), is also unlikely since the target syllables prompting repetitions remained constant across conditions. Our findings indicate that even in the absence of interactive or socially driven influences, subtle acoustic context alone can modulate speech production. Moreover, finding that nonspeech tone contexts entirely devoid of articulatory-phonetic information influence production is a challenge to theories emphasizing the need for detailed phonetic encoding of another talker's speech in articulatory terms (e.g., Pickering & Garrod, 2013; Shockley et al., 2004). None of this denies the potential for social and interactive factors to influence phenomena like phonetic convergence. Instead, our results make the case speech production adapts even in listening contexts that possess no opportunities for mimicry, acoustic-phonetic parsing, or tuning through articulatory-phonetic information. This indicates that the representational currency linking speech perception and production is not exclusively phonetic. Speech motor control is affected by general auditory processes.

This poses both challenges and opportunities for speech motor planning models, which have been built to understand how sensory feedback from *one's own speech* refines predictive feedforward motor control (Guenther et al., 2006; Hickok, 2012; Houde & Chang, 2015;

Houde & Nagarajan, 2011; Parrell et al., 2019; Parrell & Houde, 2019; Tian & Poeppel, 2010). Although these models do not specifically address perception-production interactions driven by "external" speech such as phonetic convergence, extensions of these models might parsimoniously account for influences of other talkers' speech. Toward this goal, the present study informs our understanding of perception-production interactions in several ways.

Firstly, models of language production posit stable representations that are simply "accessed." This is true even for tasks that involve repeating nonsense syllables spoken by an interlocutor (e.g., Dell et al., 2013; Nozari et al., 2010; Nozari & Dell, 2013). This is also true of many models of speech motor control, even though they have a more sophisticated articulatory-phonetic space (Guenther, 2016; Houde & Nagarajan, 2011). However, individual differences in speech perception have been long recognized to relate to production. The more accurately a speaker discriminates a speech contrast, the more distinctly the speaker produces the contrast (Perkell et al., 2004, 2014), for example. Our within-participant examination of listening contexts that selectively enhance and diminish the perceptual distinctiveness of target syllables illustrates that this relationship extends beyond stable individual differences to reflect online tuning of perceptual processing across contexts, even *within individual listeners.* This argues that the speech representations that define the goals of speech production in multidimensional auditory space (Guenther, 1995) are more malleable and subject to short-term perceptual influence of other voices – and indeed even nonspeech sounds – than has been typically modelled.

Secondly, the results speak to the locus of these malleable representations. Bourguignon et al. (2016) found that (spectrally contrastive) shifts in vowel perception due to a precursor phrase influence sensorimotor adaptation under altered auditory feedback. External speech thus influences how auditory feedback from one's own voice impacts speech motor control. Bourguignon and colleagues speculate that external speech may influence early perceptual representation of speech targets or, alternatively, shift the predicted auditory outcome of speech production. The influence of nonspeech contexts on speech production in Experiment 3 aligns best with an early perceptual locus.

This possibility aligns with conclusions emerging from the influence of statistical learning across accented speech on speech production (Murphy, Holt, & Nozari, 2025; Murphy, Nozari, & Holt, 2025). A closer examination of the perceptual statistical learning that induces these changes in speech production reveals that exposure to an accent initiates a cascade of rapid, yet lasting, neural adjustments that first emerge at short cortical latencies associated with encoding sound features in auditory cortex (Llanos et al., 2025). These perceptual effects carry over to influence production (Murphy, Holt & Nozari, 2025; Murphy, Nozari & Holt., 2025), suggesting that perceptual tuning of the auditory cortical encoding of speech influences speech production. The results of Experiment 3, especially, align with this possibility. General auditory interactions common to speech and nonspeech are sufficient to shift speech production, ruling out the necessity of strictly articulatory-phonetic or speech-specific representations in defining the goals of speech production and arguing for an early locus of these goals in auditory cortex.

Thirdly, our study speaks to the mechanisms that drive adjustments to production. The Directions into Velocities of Articulators (DIVA) model describes an elegant and neurobiologically plausible model of speech motor control (e.g., Tourville et al., 2008). In the DIVA model, adjustments to production require overt production because it produces sensory feedback, as in hearing one's own speech. A mismatch between the expected auditory consequence of a production and the auditory feedback that arrives, such as in altered auditory feedback paradigms, drives sensorimotor adjustments to production (Guenther, 2016; Meier & Guenther, 2023). The results of Bourguignon et al. (2016), reviewed above, and other studies that demonstrate an effect of external speech on sensorimotor adaptation to altered auditory feedback (Lametti et al., 2014; Shiller & Rochon, 2014) are compatible with this mechanism because auditory sensory feedback is directly manipulated, and effects tend to emerge across multiple productions.

Recently, we have suggested extending this framework to explain changes to a listener's speech as a function of perceptual input from other people's speech (Murphy, Nozari, & Holt, 2025a), and presented evidence that these effects are evident in the very first production after a speech regularity that shifts perception, before experiencing auditory feedback (Murphy, Holt, & Nozari, 2025a). This precludes comparison of overt auditory feedback against the anticipated auditory consequences of that production in DIVA's "auditory target map." This demonstration mirrors two general advances in theories of monitoring and production; first, that internal states that are predictive of performance outcomes are sufficient for monitoring in the absence of overt production and its sensory consequences (Nozari, 2025; Runnqvist & Kell, 2025), and second, that perception-production systems are in constant flux due to incremental learning processes that create subtle but influential shifts in the representational space (Dell et al., 2021; Nozari, 2025). In the same vein, we propose that the auditory target map is susceptible to changes as a function of the statistics of incoming speech, or as in the present studies, spectral properties of the listening context. This reshapes the map directly and thus changes the auditory target of production. Here, we take this proposal one step further by shedding further light on the nature of the representations in the auditory target map. We demonstrate that the auditory target map does not simply store articulatory phonetic features of speech.

On this point, it is important to note that not all shifts in speech perception are mirrored in production. As an example, exposure to acoustic regularities that convey an accent across *bear* and *pear* utterances leads listeners to down-weight acoustic speech dimensions that violate the norm and listeners' own speech exhibits reduced distinctiveness on this dimension, as well (Murphy, Nozari & Holt, 2025b). Interestingly, the perceptual down-weighting evoked by experience across *bear-pear* generalizes to influence perception, but not production, of *beer-pier*. Thus, not all perceptual shifts alter speech motor control. The evidence presently available argues that transfer from perception to production is quite narrowly tuned.

Intriguingly, this would seem to stand in contrast to the "far" transfer observed for nonspeech listening contexts' influence on speech production in Experiment 3. The resolution of this apparent tension is likely to be related to the precise nature of the perceptual influence elicited across different tasks. Statistical learning across an accent,

explicit training with feedback, and spectrally contrastive context effects each gives rise to changes in speech production. Crucially, each is elicited by distinct perceptual processes, the nature of which is likely to influence perception-production interactions in nuanced ways. Evidence of transfer across each begins to create a toolbox of empirical phenomena with which future studies can better delineate how perceptual shifts mediate adjustments to feedforward internal models for speech motor control.

In summary, our data emphasize that the sensorimotor processes guiding speech production are not sequestered from general auditory perceptual processing. Seemingly 'low level' perceptual processes that sharpen spectral contrast across successive sounds exert an influence on the speech motor control system guiding speech production. These results highlight opportunities to extend and refine models of speech motor control to include influences of external speech, and even nonspeech, perceptual processing and to consider the representational structure of internal perceptual models that drive feedforward adjustments to speech production.

Constraints on Generality

We believe our samples to be representative of a broader population of young adult American English listeners with typical hearing, as confirmed by our replication of the perceptual spectral contrast effect that has been observed across many samples young adult samples spanning 2005 to 2025, in online studies utilizing participants' own equipment as well as in laboratory studies with highly specialized equipment, and with both experimenter-delivered instructions and automated online protocols. The English-language speech stimuli of the study directed us to an American English sample, but the conclusions of the study should generalize to other languages, especially because our data point to general auditory representations. The stimuli that elicited the effects reported here spanned isolated synthesized syllables, natural speech in the form of an audiobook, and pure tones. Replications of the perceptual effect, and its transfer to production, across this heterogeneity of stimuli assure us that the context sounds' distributions of spectral energy produce shifts in perception that transfer to production; we expect future replications attentive to these spectral distributions to elicit similar effects across new samples of listeners and new classes of stimuli. Finally, we note that the effects were robust to the change in task demands (categorization before production) introduced in Experiment 3.

Acknowledgments

This work was supported by the National Science Foundation (BCS-2346989 to NN and LLH; BCS- 2420979 to LLH and FD) and the National Institutes of Health (R01DC017734 to LLH and FD). Data, code, and stimuli can be found at https://osf.io/mnj85/?view_only=bfc195e950b5456bb93b68ce6ea03ba9 (view-only link for review). Some of the data from this study were presented at the 65th Annual Meeting of the Psychonomic Society. The authors have no known conflict of interest to disclose and thank Professor Alessandro Rinaldo for statistical consultation.

References

Anwyl-Irvine AL, Massonnié J, Flitton A, Kirkham N, & Evershed JK (2020). Gorilla in our midst: An online behavioral experiment builder. Behavior Research Methods, 52(1), 388–407. 10.3758/s13428-019-01237-x [PubMed: 31016684]

Babel M (2010). Dialect divergence and convergence in New Zealand English. Language in Society, 39(4), 437–456.

Babel M (2012). Evidence for phonetic and social selectivity in spontaneous phonetic imitation. Journal of Phonetics, 40(1), 177–189. 10.1016/j.wocn.2011.09.001

Bartlett EL, & Wang X (2005). Long-lasting modulation by stimulus context in primate auditory cortex. Journal of Neurophysiology, 94(1), 83–104. 10.1152/jn.01124.2004 [PubMed: 15772236]

Bates D, Mächler M, Bolker B, & Walker S (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67, 1–48. 10.18637/jss.v067.i01

Boersma P, & Weenink D (2025). Praat: Doing phonetics by computer (Version Version 6.4.28) [Computer software].

Bosker HR, Sjerps MJ, & Reinisch E (2020). Spectral contrast effects are modulated by selective attention in “cocktail party” settings. Attention, Perception, & Psychophysics, 82(3), 1318–1332. 10.3758/s13414-019-01824-2

Bourguignon NJ, Baum SR, & Shiller DM (2016). Please say what this word is—Vowel-extrinsic normalization in the sensorimotor control of speech. Journal of Experimental Psychology: Human Perception and Performance, 42(7), 1039–1047. 10.1037/xhp0000209 [PubMed: 26820250]

Bourhis RY, & Giles H (1977). The Language of Intergroup Distinctiveness. In Language Ethnicity and Intergroup Relations (pp. 119–135). Academic Press.

Bradshaw AR, Lametti DR, Shiller DM, Jasmin K, Huang R, & McGettigan C (2023). Speech motor adaptation during synchronous and metronome-timed speech. Journal of Experimental Psychology: General, 152(12), 3476–3489. 10.1037/xge0001459 [PubMed: 37616075]

Degras D (2017). Simultaneous confidence bands for the mean of functional data. WIREs Computational Statistics, 9(3), e1397. 10.1002/wics.1397

Delattre PC, Liberman AM, & Cooper FS (1955). Acoustic loci and transitional cues for consonants. Journal of the Acoustical Society of America, 27, 769–773. 10.1121/1.1908024

Dell GS, Kelley AC, Hwang S, & Bian Y (2021). The adaptable speaker: A theory of implicit learning in language production. Psychological Review, 128(3), 446. [PubMed: 33705201]

Dell GS, Schwartz MF, Nozari N, Faseyitan O, & Branch Coslett H (2013). Voxel-based lesion-parameter mapping: Identifying the neural correlates of a computational model of word production. Cognition, 128(3), 380–396. 10.1016/j.cognition.2013.05.007 [PubMed: 23765000]

Earnshaw K (2021). Examining the implications of speech accommodation for forensic speaker comparison casework: A case study of the West Yorkshire face vowel. Journal of Phonetics, 87, 101062. 10.1016/j.wocn.2021.101062

Fowler CA (2006). Compensation for coarticulation reflects gesture perception, not spectral contrast. Perception & Psychophysics, 68(2), 161–177. 10.3758/BF03193666 [PubMed: 16773890]

Frossard J, & Renaud O (2021). Permutation Tests for Regression, ANOVA, and Comparison of Signals: The permuco Package. Journal of Statistical Software, 99, 1–32. 10.18637/jss.v099.i15

Giles H, Coupland N, & Coupland J (1991). Accommodation theory: Communication, context, and consequence. In Giles H, Coupland J, & Coupland N (Eds.), Contexts of Accommodation: Developments in Applied Sociolinguistics (pp. 1–68). Cambridge University Press. 10.1017/CBO9780511663673.001

Giles H, Edwards AL, & Walther JB (2023). Communication accommodation theory: Past accomplishments, current trends, and future prospects. Language Sciences, 99, 101571. 10.1016/j.langsci.2023.101571

Gregory SW, & Webster S (1996). A nonverbal signal in voices of interview partners effectively predicts communication accommodation and social status perceptions. Journal of Personality and Social Psychology, 70(6), 1231–1240. 10.1037/0022-3514.70.6.1231 [PubMed: 8667163]

Guenther FH (1995). Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production. Psychological Review, 102(3), 594–621. 10.1037/0033-295X.102.3.594 [PubMed: 7624456]

Guenther FH (2016). Neural Control of Speech. https://direct.mit.edu/books/monograph/4085/Neural-Control-of-Speech

Guenther FH, Ghosh SS, & Tourville JA (2006). Neural modeling and imaging of the cortical interactions underlying syllable production. Brain and Language, 96(3), 280–301. 10.1016/j.bandl.2005.06.001 [PubMed: 16040108]

Heath J (2015). Convergence through divergence: Compensatory changes in phonetic accommodation. LSA Annual Meeting Extended Abstracts, 6, 7:1–5. 10.3765/exabs.v0i0.3002

Hickok G (2012). Computational neuroanatomy of speech production. Nature Reviews Neuroscience, 13(2), 135–145. 10.1038/nrn3158 [PubMed: 22218206]

Hodson AJ, Shinn-Cunningham BG, & Holt LL (2023). Statistical learning across passive listening adjusts perceptual weights of speech input dimensions. Cognition, 238, 105473. 10.1016/j.cognition.2023.105473 [PubMed: 37210878]

Holt LL (2005). Temporally Nonadjacent Nonlinguistic Sounds Affect Speech Categorization. Psychological Science, 16(4), 305–312. 10.1111/j.0956-7976.2005.01532.x [PubMed: 15828978]

Holt LL (2006a). Speech categorization in context: Joint effects of nonspeech and speech precursors. The Journal of the Acoustical Society of America, 119(6), 4016–4026. 10.1121/1.2195119 [PubMed: 16838544]

Holt LL (2006b). The mean matters: Effects of statistically defined nonspeech spectral distributions on speech categorization. The Journal of the Acoustical Society of America, 120(5), 2801–2817. 10.1121/1.2354071 [PubMed: 17091133]

Holt LL, & Lotto AJ (2002). Behavioral examinations of the level of auditory processing of speech context effects. Hearing Research, 167(1), 156–169. 10.1016/S0378-5955(02)00383-0 [PubMed: 12117538]

Holt LL, & Lotto AJ (2006). Cue weighting in auditory categorization: Implications for first and second language acquisition. The Journal of the Acoustical Society of America, 119(5 Pt 1), 3059–3071. 10.1121/1.2188377 [PubMed: 16708961]

Holt LL, Lotto AJ, & Kluender KR (2000). Neighboring spectral content influences vowel identification. Journal of the Acoustical Society of America, 108(2), 710–722. 10.1121/1.429604 [PubMed: 10955638]

Houde JF, & Chang EF (2015). The cortical computations underlying feedback control in vocal production. Current Opinion in Neurobiology, 33, 174–181. 10.1016/j.conb.2015.04.006 [PubMed: 25989242]

Houde JF, & Jordan MI (1998). Sensorimotor adaptation in speech production. Science (New York, N.Y.), 279(5354), 1213–1216. 10.1126/science.279.5354.1213 [PubMed: 9469813]

Houde JF, & Nagarajan SS (2011). Speech production as state feedback control. Frontiers in Human Neuroscience, 5, 82. 10.3389/fnhum.2011.00082 [PubMed: 22046152]

Huang J, & Holt LL (2012). Listening for the Norm: Adaptive Coding in Speech Categorization. Frontiers in Psychology, 3. https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2012.00010

Idemaru K, & Holt LL (2011). Word Recognition Reflects Dimension-based Statistical Learning. Journal of Experimental Psychology. Human Perception and Performance, 37(6), 1939–1956. 10.1037/a0025641 [PubMed: 22004192]

Klatt DH (1980). Software for a cascade/parallel formant synthesizer. The Journal of the Acoustical Society of America, 67(3), 971–995. 10.1121/1.383940

Ladefoged P, & Broadbent DE (1957). Information conveyed by vowels. Journal of the Acoustical Society of America, 29, 98–104. 10.1121/1.1908694

Laing EJ, Liu R, Lotto AJ, & Holt LL (2012). Tuned with a Tune: Talker Normalization via General Auditory Processes. Frontiers in Psychology, 3. https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2012.00203

Lametti DR, Krol SA, Shiller DM, & Ostry DJ (2014). Brief periods of auditory perceptual training can determine the sensory targets of speech motor learning. Psychological Science, 25(7), 1325–1336. [PubMed: 24815610]

Lametti DR, Nasir SM, and Ostry DJ (2012). Sensory preference in speech production revealed by simultaneous alteration of auditory and somatosensory feedback. Journal of Neuroscience, 32, 9351–9358. 10.1523/JNEUROSCI.0404-12.2012.570 [PubMed: 22764242]

Larson CR, Altman KW, Liu H, and Hain TC (2008). Interactions between auditory and somatosensory feedback for voice F0 control. Experimental Brain Research, 187, 613–621. 10.1007/s00221-008-5721330-z. [PubMed: 18340440]

Liberman AM; Cooper FS; Shankweiler DP; Studdert-Kennedy M (1967). Perception of the speech code. Psychological Review, 7 (6), 431–461.

Lindblom BEF, & Studdert-Kennedy M (1967). On the role of formant transitions in vowel recognition. Journal of the Acoustical Society of America, 42(4), 830–843. 10.1121/1.1910655 [PubMed: 6075568]

Lindsay S, Clayards Meghan, Gennari Silvia, & and Gaskell MG (2022). Plasticity of categories in speech perception and production. Language, Cognition and Neuroscience, 37(6), 707–731. 10.1080/23273798.2021.2018471

Llanos F, Yunan W, Abel TJ, & Holt L (2025). Rapid neurodynamicc adjustments across multiple levels of language processing drive perceptual adaptation to accented speech. Manuscript submitted for publication.

Lotto AJ, & Kluender KR (1998). General contrast effects in speech perception: Effect of preceding liquid on stop consonant identification. Perception & Psychophysics, 60(4), 602–619. 10.3758/BF03206049 [PubMed: 9628993]

Lotto AJ, Kluender KR, & Holt LL (1997). Perceptual compensation for coarticulation by Japanese quail (Coturnix coturnix japonica). The Journal of the Acoustical Society of America, 102(2 Pt 1), 1134–1140. 10.1121/1.419865 [PubMed: 9265760]

Lotto AJ, Sullivan SC, & Holt LL (2003). Central locus for nonspeech context effects on phonetic identification (L). The Journal of the Acoustical Society of America, 113(1), 53–56. 10.1121/1.1527959 [PubMed: 12558245]

Mann VA (1980). Influence of preceding liquid on stop-consonant perception. Perception & Psychophysics, 28(5), 407–412. 10.3758/BF03204884 [PubMed: 7208250]

Mann VA (1986). Distinguishing universal and language-dependent levels of speech perception: Evidence from Japanese listeners' perception of English "l" and "r." Cognition, 24(3), 169–196. 10.1016/S0010-0277(86)80001-4 [PubMed: 3816123]

Mann VA, & Repp BH (1980). Influence of vocalic context on the perception of the [ʃ]-[s] distinction. Perception and Psychophysics, 28, 213–228. [PubMed: 7432999]

Mann VA, & Repp BH (1981). Influence of preceding fricative on stop consonant perception. Journal of the Acoustical Society of America, 69, 548–558. [PubMed: 7462477]

Maris E, & Oostenveld R (2007). Nonparametric statistical testing of EEG- and MEG-data. Journal of Neuroscience Methods, 164(1), 177–190. 10.1016/j.jneumeth.2007.03.024 [PubMed: 17517438]

Maye J, Weiss DJ, & Aslin RN (2008). Statistical phonetic learning in infants: Facilitation and feature generalization. Developmental Science, 11(1), 122–134. 10.1111/j.1467-7687.2007.00653.x [PubMed: 18171374]

Meier AM, & Guenther FH (2023). Neurocomputational modeling of speech motor development. Journal of Child Language, 50(6), 1318–1335. 10.1017/S0305000923000260 [PubMed: 37337871]

Michalsky J, & Schoormann H (2017). Pitch Convergence as an Effect of Perceived Attractiveness and Likability. 2253–2256. 10.21437/Interspeech.2017-1520

Miller RM, Sanchez K, & Rosenblum LD (2010). Alignment to visual speech information. Attention, Perception, & Psychophysics, 72(6), 1614–1625. 10.3758/APP.72.6.1614

Milne AE, Bianco R, Poole KC, Zhao S, Oxenham AJ, Billig AJ, & Chait M (2021). An online headphone screening test based on dichotic pitch. Behavior Research Methods, 53(4), 1551–1562. 10.3758/s13428-020-01514-0 [PubMed: 33300103]

Mukherjee S, D'Ausilio A, Nguyen N, Fadiga L, & Badino L (2017). The Relationship Between F0 Synchrony and Speech Convergence in Dyadic Interaction. 2341–2345. 10.21437/Interspeech.2017-795

Murphy T, Holt LL, & Nozari N (2025b). Exposure to an Accent Transfers to Speech Production in a Single Shot. OSF. 10.31234/osf.io/qwy9v_v2

Murphy T, Nozari N, & Holt LL (2024). Transfer of statistical learning from passive speech perception to speech production. Psychonomic Bulletin & Review. 10.3758/s13423-023-02399-8

Murphy T, Nozari N, & Holt LL (2025a). Bears don't always mess with beers: Limits on generalization of statistical learning in speech. Psychonomic Bulletin & Review. 10.3758/s13423-025-02690-w

Namy LL, Nygaard LC, & Sauerteig D (2002). Gender Differences in Vocal Accommodation: The Role of Perception. Journal of Language and Social Psychology, 21(4), 422–432. 10.1177/026192702237958

Nielsen K (2011). Specificity and abstractness of VOT imitation. Journal of Phonetics, 39(2), 132–142. 10.1016/j.wocn.2010.12.007

Nozari N, & Dell GS (2013). How damaged brains repeat words: A computational approach. Brain and Language, 126(3), 327–337. 10.1016/j.bandl.2013.07.005 [PubMed: 23933472]

Nozari N, Kittredge AK, Dell GS, & Schwartz MF (2010). Naming and repetition in aphasia: Steps, routes, and frequency effects. Journal of Memory and Language, 63(4), 541–559. 10.1016/j.jml.2010.08.001 [PubMed: 21076661]

Nozari N (2025). Monitoring, control and repair in word production. Nature Reviews Psychology, 4(3), 222–238.

Ostrand R, & Chodroff E (2021). It's alignment all the way down, but not all the way up: Speakers align on some features but not others within a dialogue. Journal of Phonetics, 88, 101074. 10.1016/j.wocn.2021.101074 [PubMed: 34366499]

Pardo JS (2006). On phonetic convergence during conversational interaction. The Journal of the Acoustical Society of America, 119(4), 2382–2393. [PubMed: 16642851]

Pardo JS, Jay IC, Hoshino R, Hasbun SM, Sowemimo-Coker C, & Krauss RM (2013). Influence of Role-Switching on Phonetic Convergence in Conversation. Discourse Processes, 50(4), 276–300. 10.1080/0163853X.2013.778168

Pardo JS, Jay IC, & Krauss RM (2010). Conversational role influences speech imitation. Attention, Perception, & Psychophysics, 72(8), 2254–2264. 10.3758/BF03196699

Pardo JS, Pellegrino E, Dellwo V, & Möbius B (2022). Vocal accommodation in speech communication. Journal of Phonetics, 95, 101196.

Pardo JS, Urmanche A, Wilman S, & Wiener J (2017). Phonetic convergence across multiple measures and model talkers. Attention, Perception, & Psychophysics, 79(2), 637–659. 10.3758/s13414-016-1226-0

Pardo JS, Urmanche A, Wilman S, Wiener J, Mason N, Francis K, & Ward M (2018). A comparison of phonetic convergence in conversational interaction and speech shadowing. Journal of Phonetics, 69, 1–11. 10.1016/j.wocn.2018.04.001

Parrell B, & Houde J (2019). Modeling the role of sensory feedback in speech motor control and learning. Journal of Speech, Language, and Hearing Research, 62(8, Suppl), 2963–2985. 10.1044/2019_JSLHR-S-CSMC7-18-0127

Parrell B, Ramanarayanan V, Nagarajan S, & Houde J (2019). The FACTS model of speech motor control: Fusing state estimation and task-based control. PLOS Computational Biology, 15(9), e1007321. 10.1371/journal.pcbi.1007321 [PubMed: 31479444]

Perkell JS, Guenther FH, Lane H, Matthies ML, Stockmann E, Tiede M, & Zandipour M (2004). The distinctness of speakers' productions of vowel contrasts is related to their discrimination of the contrasts. The Journal of the Acoustical Society of America, 116(4), 2338–2344. 10.1121/1.1787524 [PubMed: 15532664]

Perkell JS, Matthies ML, Tiede M, Lane H, Zandipour M, Marrone N, Stockmann E, & Guenther FH (2014). The Distinctness of Speakers' /s/—/ʃ/ Contrast Is Related to Their Auditory Discrimination and Use of an Articulatory Saturation Effect. ASHA. https://pubs.asha.org/doi/full/10.1044/1092-4388(2004/095)

Pickering MJ, & Garrod S (2013). An integrated theory of language production and comprehension. Behavioral and Brain Sciences, 36(4), 329–347. 10.1017/S0140525X12001495 [PubMed: 23789620]

Purcell DW, & Munhall KG (2006). Compensation following real-time manipulation of formants in isolated vowels. The Journal of the Acoustical Society of America, 119(4), 2288–2297. 10.1121/1.2173514 [PubMed: 16642842]

Repp BH, & Mann VA (1981). Perceptual assessment of fricative—Stop coarticulation. The Journal of the Acoustical Society of America, 69(4), 1154–1163. 10.1121/1.385695 [PubMed: 7229203]

Rochet-Capellan A, & Ostry DJ (2011). Simultaneous acquisition of multiple auditory–motor transformations in speech. The Journal of Neuroscience, 31(7), 2657–2662. 10.1523/JNEUROSCI.6020-10.2011 [PubMed: 21325534]

Rowling J (1998). Harry Potter and the Sorcerer's Stone. Arthur A. Levine Books.

Runnqvist E, & Kell CA (2025). A continuum of predictive control between motor and mental actions: language production as a test case. Current Opinion in Behavioral Sciences, 65, 101573.

Sassenhagen J, & Draschkow D (2019). Cluster-based permutation tests of MEG/EEG data do not establish significance of effect latency or location. Psychophysiology, 56(6), e13335. 10.1111/psyp.13335 [PubMed: 30657176]

Sato M, Grabski K, Garnier M, Granjon L, Schwartz J-L, & Nguyen N (2013). Converging toward a common speech code: Imitative and perceptuo-motor recalibration processes in speech production. Frontiers in Psychology, 4. 10.3389/fpsyg.2013.00422

Schertz J, & Paquette-Smith M (2023). Convergence to shortened and lengthened voice onset time in an imitation task. JASA Express Letters, 3(2), 025201. 10.1121/10.0017066 [PubMed: 36858990]

Shiller DM, & Rochon M-L (2014). Auditory-perceptual learning improves speech motor adaptation in children. Journal of Experimental Psychology. Human Perception and Performance, 40(4), 1308–1315. 10.1037/a0036660 [PubMed: 24842067]

Shockley K, Sabadini L, & Fowler CA (2004). Imitation in shadowing words. Perception & Psychophysics, 66(3), 422–429. 10.3758/BF03194890 [PubMed: 15283067]

Shrem Y, Goldrick M, & Keshet J (2019). Dr.VOT: Measuring Positive and Negative Voice Onset Time in the Wild. 629–633. 10.21437/Interspeech.2019-1735

Song P, Zhai Y, & Yu X (2023). Stimulus-specific adaptation (SSA) in the auditory system: Functional relevance and underlying mechanisms. Neuroscience & Biobehavioral Reviews, 149, 105190. 10.1016/j.neubiorev.2023.105190 [PubMed: 37085022]

Stephens JDW, & Holt LL (2003). Preceding phonetic context affects perception of nonspeech (L). The Journal of the Acoustical Society of America, 114(6), 3036–3039. 10.1121/1.1627837 [PubMed: 14714784]

Stevens KN (1998). Acoustic Phonetics. MIT Press.

Stilp CE, & Assgari AA (2018). Perceptual sensitivity to spectral properties of earlier sounds during speech categorization. Attention, Perception, & Psychophysics, 80(5), 1300–1310. 10.3758/s13414-018-1488-9

The MathWorks Inc. (2024). MATLAB (Version Version 23.2.0 (R2023b)) [Computer software].

Tian X, & Poeppel D (2010). Mental imagery of speech and movement implicates the dynamics of internal forward models. Frontiers in Psychology, 1, 166. 10.3389/fpsyg.2010.00166 [PubMed: 21897822]

Tourville JA, Reilly KJ, & Guenther FH (2008). Neural mechanisms underlying auditory feedback control of speech. NeuroImage, 39(3), 1429–1443. 10.1016/j.neuroimage.2007.09.054 [PubMed: 18035557]

Ulanovsky N, Las L, Farkas D, & Nelken I (2004). Multiple Time Scales of Adaptation in Auditory Cortex Neurons. Journal of Neuroscience, 24(46), 10440–10453. 10.1523/JNEUROSCI.1905-04.2004 [PubMed: 15548659]

Villacorta VM, Perkell JS, & Guenther FH (2007). Sensorimotor adaptation to feedback perturbations of vowel acoustics and its relation to perception. The Journal of the Acoustical Society of America, 122(4), 2306–2319. 10.1121/1.2773966 [PubMed: 17902866]

Wade T, & Holt LL (2005). Effects of later-occurring nonlinguistic sounds on speech categorization. The Journal of the Acoustical Society of America, 118(3 Pt 1), 1701–1710. 10.1121/1.1984839 [PubMed: 16240828]

Watkins A (1991). Central, auditory mechanisms of perceptual compensation for spectral-envelope distortiona). The Journal of the Acoustical Society of America, 90(6), 2942–2955. 10.1121/1.401769 [PubMed: 1787236]

Watkins A, & Makin S (2007). Perceptual Compensation for Reverberation: Effects of 'Noise-Like' and 'Tonal' Contexts. In Kollmeier B, Klump G, Hohmann V, Langemann U, Mauermann M,

Uppenkamp S, & Verhey J (Eds.), Hearing – From Sensory Processing to Perception (pp. 533–540). Springer. 10.1007/978-3-540-73009-5_57

Public Significance Statement

What we hear influences how we talk. We examine the interplay of perception and production by asking whether different listening contexts that shift speech perception also affect speech production. Listening contexts composed of simple syllables, continuous speech, and even sequences of nonspeech tones all have an impact on speech production. This demonstrates a role for general auditory processes in defining the motor goals of speech production.

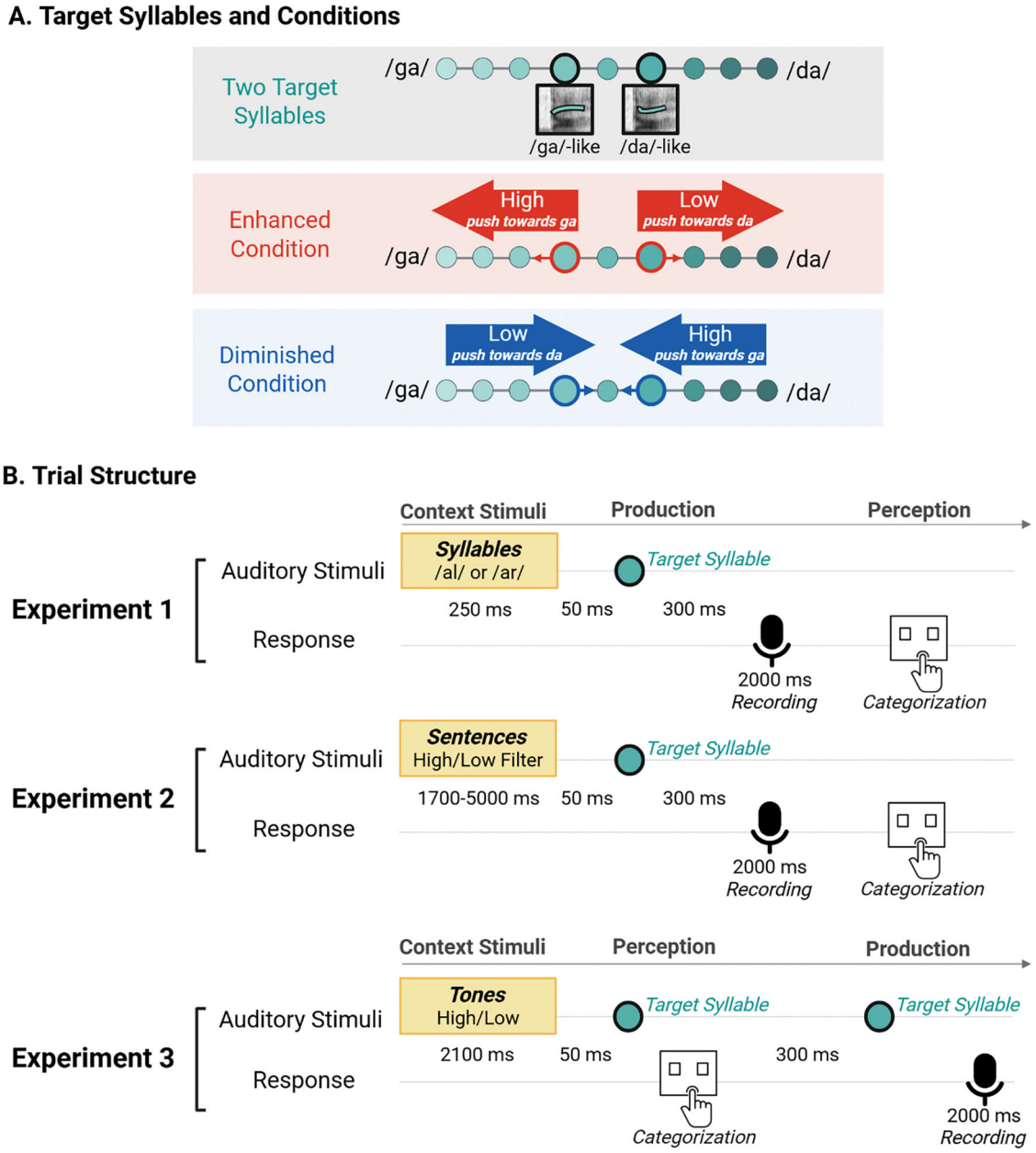

Figure 1. Experimental Design.

(A) Illustration of the approach to strategically pairing contexts with two perceptually ambiguous target syllables to create Enhanced and Diminished perceptual effects. Contexts with a concentration of High versus Low spectral energy are paired with two perceptually ambiguous target syllables (syllables 4 and 6 on a 9-point /ga/-/da/ series), one more acoustically /ga/-like and one more /da/-like. **(B)** Trial structure for each experiment.

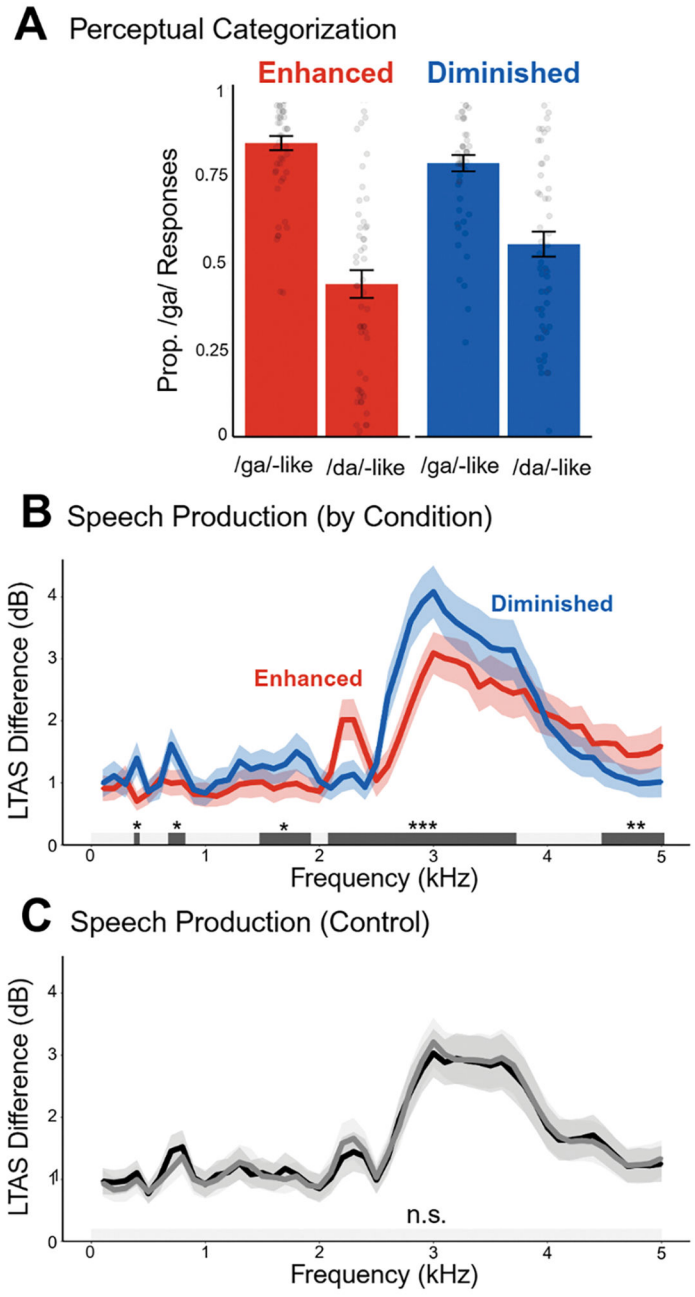

Figure 2. Results of Experiment 1.

(A) Perceptual categorization of the target syllables was influenced by the preceding context syllables in a spectrally contrastive manner. **(B)** Speech production differed as a function of whether the context syllables Enhanced or Diminished the perceptual distinctiveness of the target syllables. **(C)** A control analysis examined the LTAS Differences calculated across 100 /ga/- and 100 /da/-response trials without regard to listening context. The black line shows one random sample; the grey line shows a second random sample. Their overlap indicates that the differences in speech acoustics apparent in (B) does not arise simply from

random speech variability. Confidence bands indicate 95% of all LTAS curves across all frequencies, estimated using functional data analysis.

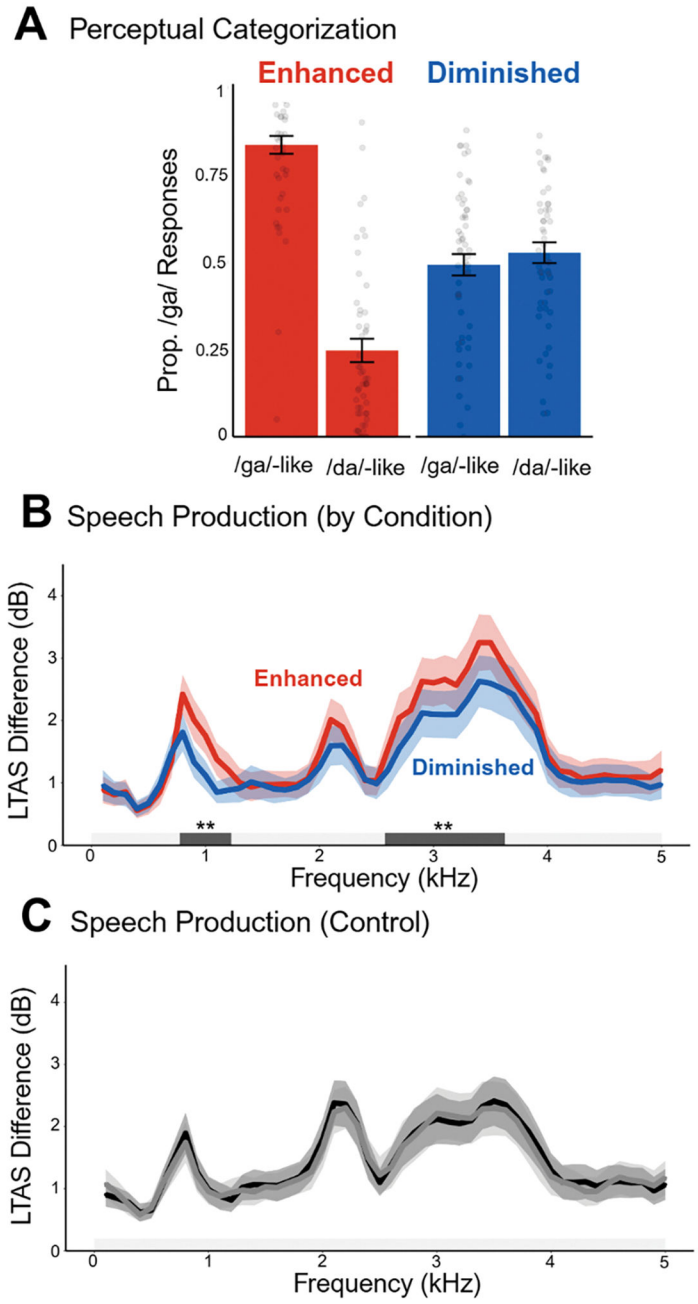

Figure 3. Results of Experiment 2.

(A) Perceptual Categorization of the target syllables was influenced by the preceding audiobook excerpt in a spectrally contrastive manner. **(B)** Speech Production differed as a function of whether the context syllables Enhanced or Diminished the perceptual distinctiveness of the target syllables. **(C)** A control analysis examined the LTAS Differences calculated across 100 /ga/- and 100 /da/-response trials without regard to listening context. The black line shows one random sample; the grey line shows a second random sample. Their overlap indicates that the differences in speech acoustics apparent in (B) does not arise

simply from random speech variability. Confidence bands indicate 95% of all LTAS curves across all frequencies, estimated using functional data analysis.

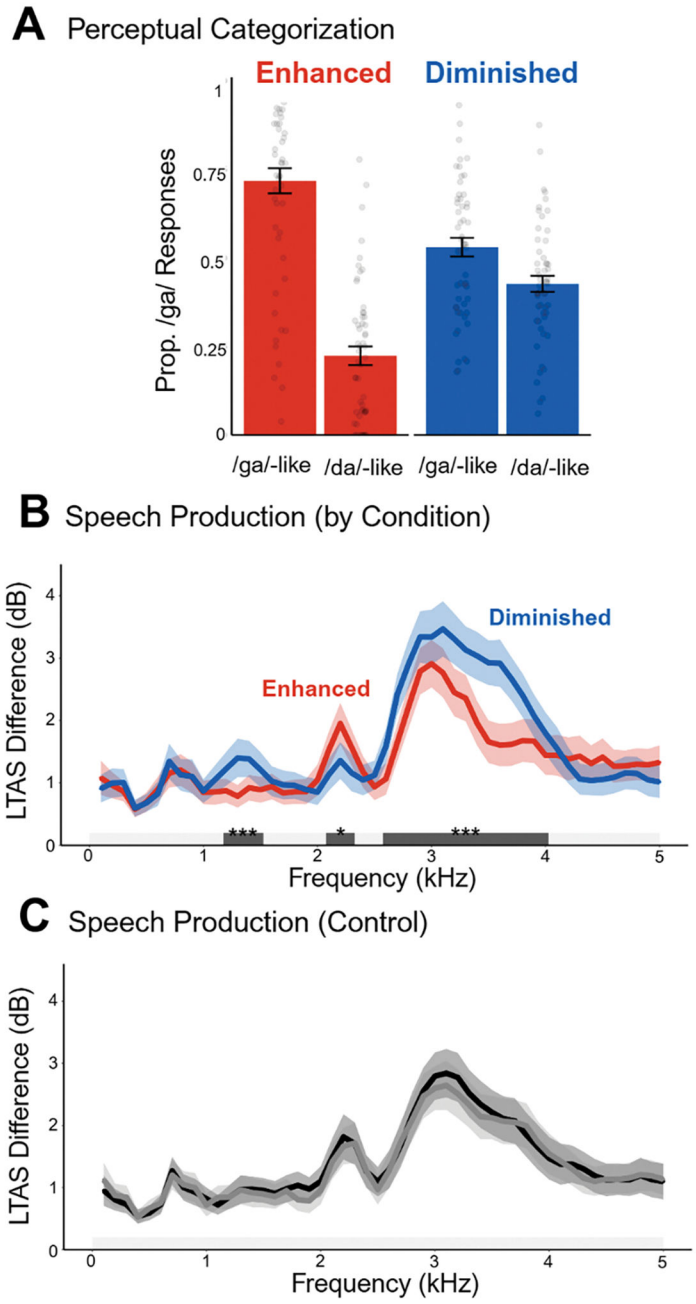

Figure 4. Results of Experiment 3.

(A) Perceptual categorization of the target syllables was influenced by the preceding tone sequences in a spectrally contrastive manner. **(B)** Speech production differed as a function of whether the context syllables Enhanced or Diminished the perceptual distinctiveness of the target syllables. **(C)** A control analysis examined the LTAS Differences calculated across 100 /ga/- and 100 /da/-response trials without regard to listening context. The black line shows one random sample; the grey line shows a second random sample. Their overlap indicates that the differences in speech acoustics apparent in (B) does not arise simply from

random speech variability. Confidence bands indicate 95% of all LTAS curves across all frequencies, estimated using functional data analysis.

Table I.

Experiment 1 Perceptual Categorization

	Estimate (β)	Std. Error	z	p
(Intercept)	1.52	0.14	10.98	< .001
Condition	0.42	0.07	6.18	< .001
Target Syllable	−1.21	0.06	−20.45	< .001
Condition × Syllable	−0.96	0.09	−11.08	< .001

Table II.

Experiment 1 Cluster-based Permutation Tests of Speech Production LTAS Difference across Enhanced and Diminished Conditions

Cluster Region	Cluster Mass (t)	p
400 Hz	50.38	.019
700–800 Hz	39.57	.034
1,500–1,900 Hz	55.99	.016
2,100–3,700 Hz	466.05	< .001
4,500–5,000 Hz	87.23	.005

Table III.

Experiment 2 Perceptual Categorization

	Estimate (β)	Std. Error	z	p
(Intercept)	−0.03	0.13	−0.26	0.797
Condition	1.93	0.07	29.27	< .001
Target Syllable	0.17	0.06	3.03	0.002
Condition × Syllable	−3.38	0.09	−37.3567	< .001

Table IV.

Experiment 2 Cluster-based Permutation Tests of Speech Production LTAS Difference across Enhanced and Diminished Conditions

Cluster Region	Cluster Mass (t)	p
800–1,200 Hz	107.12	0.002
2,600–3,600 Hz	86.99	0.005

Table V.

Experiment 3 Perceptual Categorization

	Estimate (β)	Std. Error	z	p
(Intercept)	0.13	0.09	1.43	0.152
Condition	0.87	0.078	11.26	< .001
Target Syllable	−0.45	0.07	−6.13	< .001
Condition × Syllable	−1.89	0.11	−16.88	< .001

Table VI.

Experiment 3 Cluster-based Permutation Tests of Speech Production LTAS Difference across Enhanced and Diminished Conditions

Cluster Region	Cluster Mass (t)	p
1,200–1,500 Hz	63.19	.009
2,100–2,300 Hz	37.46	.033
2,600–4,000 Hz	390.47	< .001