AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology: General

ISSN: 0096-3445

2025, Vol. 154, No. 6, 1681–1698
https://doi.org/10.1037/xge0001744

Simultaneous Acquisition of Multiple Auditory–Motor Transformations Reveals Suprasyllabic Motor Planning in Speech Production

Yuyu Zeng[1], Caroline A. Niziolek[1, 2], and Benjamin Parrell[1, 2]
[1] Waisman Center, University of Wisconsin–Madison
[2] Department of Communication Sciences and Disorders, University of Wisconsin–Madison

Motor planning forms a critical bridge between psycholinguistic and motoric models of word production. While syllables are often considered the core speech motor planning unit, growing evidence hints at suprasyllabic planning that may correspond to words, but firm experimental support is still lacking. We use differential adaptation to altered auditory feedback to provide novel, straightforward evidence for word-level planning. By introducing opposing perturbations to shared segmental content in near real time during speaking (e.g., raising the first vowel formant of “ped” in “pedigree” but lowering it in “pedicure,” so speakers hear something akin to “padigree” and “pidicure”), we assess whether participants can use the larger word context to separately oppose the two perturbations (i.e., by producing “pidigree” and “padicure”). Critically, limb control research shows that such differential learning is possible only when the shared movement forms part of distinct motor plans, allowing a straightforward assay of the scope of planning in multisyllabic words. We found differential adaptation in multisyllabic words but of smaller size relative to monosyllabic words. Our results strongly suggest that speech relies on an interactive motor planning process encompassing both syllables and words.

Public Significance Statement

This study provides strong evidence that speech motor planning involves multisyllabic units, potentially corresponding to whole words. These findings are incompatible with speech production models that assume a simple concatenation of syllables. Our results suggest a direct influence of high-level linguistic form (i.e., word) on motor planning and speech output, calling for an updated view of motor planning units in speech.

Keywords: word production, speech motor planning, syllabic planning, altered auditory feedback, sensorimotor adaptation

Supplemental materials: https://doi.org/10.1037/xge0001744.supp

Motor planning is a critical bridge between psycholinguistic and motor control models of speech production. Psycholinguistic models typically end with selecting motor plans (see Levelt, 1999, for an overview), and motor speech models explain how these selected motor plans are translated into movements and acoustic signals (see Parrell et al., 2019, for an overview). Broadly, planning in speech production includes both an early stage of psycholinguistic preparation and using the output of this early stage for articulation in the late stage of speech motor control; the present study focuses on the nature of the motor plan that bridges these

---

Joseph Toscano served as action editor.

Yuyu Zeng https://orcid.org/0000-0002-4014-3992

Caroline A. Niziolek https://orcid.org/0000-0002-6085-1371

Benjamin Parrell https://orcid.org/0000-0003-2610-2884

Caroline A. Niziolek and Benjamin Parrell shared equal authorship. All data are publicly available on the Open Science Framework at https://osf.io/ytwqp/. This work was presented at the International Workshop on Language Production (2022), the 8th International Conference on Speech Motor Control (2022), the Conference on Motor Speech (2024), and the 13th International Seminar on Speech Production (2024). This study was supported by the Division of Behavioral and Cognitive Sciences of the National Science Foundation (NSF; Grant 2120506) awarded to Caroline A. Niziolek and Benjamin Parrell.

Yuyu Zeng played a lead role in data curation, formal analysis, investigation, visualization, and writing–original draft and a supporting role in methodology, supervision, and writing–review and editing. Caroline A. Niziolek played a supporting role in formal analysis, investigation, and visualization and an equal role in conceptualization, funding acquisition, methodology, project administration, resources, supervision, and writing–review and editing. Benjamin Parrell played a supporting role in formal analysis, investigation, and visualization and an equal role in funding acquisition, methodology, project administration, resources, supervision, and writing–review and editing.

Correspondence concerning this article should be addressed to Yuyu Zeng, Waisman Center, University of Wisconsin–Madison, 1500 Highland Avenue, Room 483, Madison, WI 53705, United States. Email: yzengae@gmail.com

two stages. Both psycholinguistic and motoric models frequently position syllables as the central motor planning unit; one oft-cited example shows that high-frequency syllables are produced faster than their low-frequency counterparts (Cholin et al., 2006, 2011). However, there is also evidence that motor plans may alternatively include sub- and suprasyllabic units. Analyses of speech errors perceptually (e.g., *mell wade* for *well made*, involving sound exchanges; Shattuck-Hufnagel & Klatt, 1979) and articulatorily (e.g., coproducing /t/ and /k/ in the onset of *cop top*; Pouplier, 2007) imply planning subsyllabically at a phonemic/gestural level. By contrast, anticipatory coarticulation can cross syllable boundaries (e.g., lip rounding in anticipation of a rounded vowel /u/ in *lee scoot*; Perkell & Matthies, 1992), suggesting planning across multiple syllables.

Despite its foundational role in psycholinguistic and motoric models of speech production, our current understanding of the units and scope of speech motor planning is incomplete. Existing evidence often relies on relatively indirect measures that suffer from interpretive ambiguities (e.g., speech errors, reaction times, and timing differences). For example, the aforementioned phoneme exchange error could be attributed to early planning or late motor execution; similar concerns (i.e., inability to determine whether the locus of an observed effect is at early planning or late execution) also apply to findings based on reaction times and timing differences. An alternative method can separate psycholinguistic planning from motor execution, thus, more directly assessing the scope of motor planning, is needed.

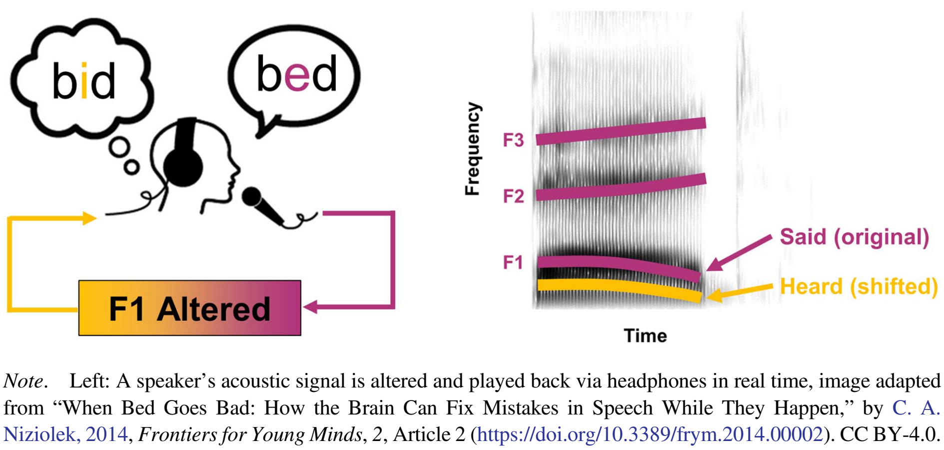
The sensorimotor adaptation procedure, which evokes learned changes to movement plans, is well-suited for this task. In speech, this procedure applies real-time perturbations to speakers' formants (the primary acoustic correlates of vowels; F1, F2, and F3 in Figure 1), so they hear a different vowel quality (e.g., "bed" is heard more like "bad"). Over repeated exposures, speakers adapt their speech to counteract the auditory perturbation (e.g., by changing the production of "bed" to be more like "bid"). This learning transfers to untrained syllables sharing the trained vowel (e.g., training on "bé" transfers to "pé"), offering indirect evidence for subsyllabic planning at the vowel level (Caudrelier et al., 2018).

When opposing perturbations are applied to the same movement in different contexts (e.g., increasing F1 in the "ped" of *pedigree* but decreasing F1 in the "ped" of *pedicure*), the presence of differential adaptation (decreasing F1 in the "ped" of *pedigree*, but increasing F1 in the "ped" of *pedicure* in the production) can directly inform us about the planning scope. The idea that distinct motor plans are necessary and sufficient for the emergence of differential adaptation has accumulated strong experimental support from studies of upper limb control. In these studies, participants are exposed to opposing perturbations to the same movement across trials, using dynamic force-field or visual perturbations. A large body of work has shown that arbitrary differences unrelated to movement planning, such as the color of the visual cue (e.g., applying leftward perturbation when a target is red, but rightward perturbation when this target is blue), do not enable differential adaptations to two opposing perturbations, even when they are fully predictive (e.g., Howard et al., 2013; see Heald et al., 2023, for a recent review).

However, differential adaptation does occur when the perturbed movement is immediately followed by a second, unperturbed, movement whose direction is predictive of the perturbation (e.g., leftward perturbation followed by a rightward movement; Howard et al., 2015). Crucially, when participants planned both the first and second movements, but did not execute the second movement (the second target disappeared during the first movement, indicating that participants should withhold the second movement), differential adaptation still occurred. By contrast, execution of the contextual movement cue without coplanning did not enable differential adaptation: When the second target only appeared after the first movement was initiated (thus, the contextual movement cue was planned separately from the first movement), there was no differential adaptation (Sheahan et al., 2016). These results show that differential adaptation is tied to separate motor plans rather than to different kinematic contexts. Collectively, studies of upper limb control suggest that motor planning differences are necessary and

Figure 1

The Speech Sensorimotor Adaptation Procedure



Note. Left: A speaker's acoustic signal is altered and played back via headphones in real time, image adapted from "When Bed Goes Bad: How the Brain Can Fix Mistakes in Speech While They Happen," by C. A. Niziolek, 2014, *Frontiers for Young Minds*, *2*, Article 2 (https://doi.org/10.3389/frym.2014.00002). CC BY-4.0. Right: Details of the modification of the speech signal. The darker bands are energy concentrations at specific frequencies, corresponding to vowel formants (labeled F1, F2, and F3), the primary acoustic correlates of vowels. In the example, the F1 of the word "bed" is perturbed downward to sound more like "bid." F1 = first formant; F2 = second formant; F3 = third formant. See the online article for the color version of this figure.

sufficient to drive differential adaptation under opposing perturbations. Accordingly, the presence/absence of differential learning can characterize the planning scope (i.e., whether or not a given context is part of the motor plan).

In speech, speakers exposed to opposing perturbations in "bed" and "head" adapted separately in each word (Rochet-Capellan & Ostry, 2011), suggesting that these monosyllabic words form distinct planning units; if the vowel were the only planning unit, what was learned from one word would be canceled out by the other. Nevertheless, because monosyllabic words are distinct syllables as well as distinct words, it is unclear which context enabled the differential adaptation. More problematically, Osu et al. (2004) found that in a random perturbation schedule that allowed exposure to the same perturbation on sequential trials (instead of alternating the perturbations), contextual differences unrelated to planning could drive differential adaptation in reaching, though this result has not been replicated (Howard et al., 2013). Correspondingly, the differential adaptation in Rochet-Capellan and Ostry's (2011) study might be explained by the sequential exposure to the same perturbation rather than the different word/syllable contexts.

In brief, previous research indicates phonemes (consonants and vowels) and syllables as motor planning units. However, firm experimental support for planning above the syllable (i.e., suprasyllabic planning) is still lacking. Here, we apply opposing perturbations to investigate the word as a potential suprasyllabic unit of speech motor planning. After confirming that speakers adapt to opposing perturbations to the same vowel in distinct monosyllabic words without sequential exposure to the same perturbation (Experiment 1), we introduced opposing perturbations to the same syllable in different disyllabic and trisyllabic words (Experiments 2 and 3). If speech motor planning incorporates word-level context independently of the syllable/vowel, adaptation should occur when the same syllable forms part of distinct multisyllabic words (e.g., opposing perturbations to "ped" in *pedigree* and *pedicure* would induce differential learning; Hypothesis 1). Conversely, if speech motor planning is purely syllabic, speakers will not adapt, as these words share a single syllabic plan and learning from one word will nullify the other (Hypothesis 2). Alternatively, if syllabic and word-level planning interact, speakers should still exhibit differential adaptation in the multisyllabic words, but the size of the differential adaptation should be reduced relative to monosyllabic words, due to the conflict between word- and syllabic-level planning (differential adaptation enabled by distinct words vs. canceled-out adaptation at the syllable level; Hypothesis 3).

Method

Participants

Fifteen speakers participated in Experiment 1 (14 females and one nonbinary; $M_{\text{age}} = 19.27$, $SD = 1.42$, range = [19, 23]), 20 speakers in Experiment 2 (17 females and three males; $M_{\text{age}} = 19.90$, $SD = 2.04$, range = [18, 27]), and 20 speakers in Experiment 3 (17 females and three males; $M_{\text{age}} = 21.30$, $SD = 2.91$, range = [19, 27]). The sample size for Experiment 1 followed Rochet-Capellan and Ostry's (2011) study, using the same monosyllabic stimuli as this prior study. Sample sizes for Experiments 2 and 3 were determined by a power analysis based on a recent study from our group with a comparable design ($d = 1.35$; Parrell & Niziolek, 2021). Accordingly, 20 participants yielded a power of approximately 1 to detect an effect of similar size and provided sufficient power to detect effects much smaller than those previously reported (0.80 at $\alpha = .05$ to detect an effect of $d = 0.58$). All participants were adult native English speakers, with some additionally speaking at least one other language (four in Experiment 1, seven in Experiment 2, and eight in Experiment 3). All participants had no reported history of neurological, speech, or hearing disorders and passed a Hughson–Westlake audiology screening before participation (thresholds ≤25 dB hearing level bilaterally in the 250–4000 Hz range). Participants were compensated either with course credit or monetary payment.

Procedure and Equipment

Participants were seated in front of a computer screen where one word appeared per trial. Participants were instructed to read these stimuli aloud as they appeared on the screen. A Sennheiser MKE 600 microphone recorded the speakers' productions, which were digitized using a Scarlett 2i2 sound card, processed (and in some trials altered by shifting vowel formants, see below) using Audapter (Cai et al., 2008; Tourville et al., 2013), and played back over Beyerdynamic DT 770 PRO closed-back, circumaural headphones. The recording, processing, and playback occurred in near real time (~18-ms delay; measurement following Kim et al., 2020). Speech was processed similarly through Audapter on all trials, regardless of whether a formant perturbation was applied. All stimulus presentation and data collection were done using MATLAB (MathWorks, Inc.).

On each trial, the target word appeared at the center of the screen for 1,400 ms (white text on a black background). The intertrial duration was 1,250 ms, with a 250-ms jitter. If the produced speech in a trial did not meet a prespecified intensity threshold, that trial was repeated. The intensity of speech playback varied with the intensity of the produced speech on each trial but was calibrated for each participant with a targeted average level of ~80-dB sound pressure level. The playback was mixed with speech-shaped noise at ~60-dB sound pressure level to mask air- and bone-conducted feedback.

Before the main experiment (described below), participants completed a brief calibration phase to achieve more accurate formant tracking. The words *bid*, *bet*, and *bat* were each repeated 10 times (order randomized), and the recordings were used to determine a speaker-specific linear predictive coding order, which was then used for identifying vowel formants in Audapter. Following the main experiment, participants completed a short survey of perturbation awareness and received an explanation of the study's purpose. The entire lab visit lasted ~1.5 hr.

Experiment Design

All perturbed syllables contained the same vowel /ɛ/. Experiment 1 tested adaptation to opposing F1 perturbations in three monosyllabic words: *head*, *bed*, and *ted*. One word received an upward F1 perturbation, another a downward F1 perturbation, and the third no F1 perturbation. Experiment 2 tested adaptation to opposing F1 perturbations in two disyllabic words whose initial syllables were phonemically identical: *seven* and *sever*. The initial syllable received an upward F1 perturbation in one word and a downward F1 perturbation in the other; the second syllable in both words was unperturbed. One additional unperturbed word, *level*, was included as a filler. Experiment 3 tested adaptation to opposing F1 perturbations in two trisyllabic words whose initial syllables were phonemically identical: *pedigree* and *pedicure*. As in Experiment 2, the initial syllable received an upward F1 perturbation in one word and a downward F1 perturbation in the other; the remaining syllables in both words were unperturbed. Two other unperturbed words (*pedestal* and *carbonate*) were included as fillers, but only the word *pedestal* was analyzed; this way, all three experiments were comparable. Including fillers with different initial syllables in Experiments 2 and 3 minimized the possibility that participants would preplan the first syllable before the appearance of the stimulus word. For each experiment, the assignment of formant perturbations to experimental words was roughly counterbalanced across participants (see the Supplemental Materials for details). Figure 2A shows a schematic of the design.

For all experiments, stimulus words were presented in blocks, each containing one repetition of each word. The order of words within each block was randomized. An additional constraint ensured that no two adjacent trials contained the same word across blocks, thus preventing the effects of a random perturbation schedule with sequential exposure to the same perturbation as a potential confound for any observed adaptation (cf. Osu et al., 2004). Each experiment consisted of four phases: an unperturbed baseline phase, a ramp phase during which the magnitude of F1 perturbation in the auditory feedback was progressively increased across blocks (the F1 perturbation magnitude remained the same within each block), a hold phase with a constant F1 perturbation of 125 mels, and an unperturbed washout phase. The magnitude of all perturbations was calculated in mels, a perceptual frequency scale along which equal changes in mels are perceived as equally distant across all frequencies (Stevens et al., 1937). Experiment 1 consisted of 30 baseline blocks, 30 ramp blocks, 90 hold blocks, and 30 washout blocks. Experiments 2 and 3 had 20 baseline blocks, 30 ramp blocks, 70 hold blocks, and 10 washout blocks (Figure 2B). In all experiments, a self-timed break was included every 10 blocks.

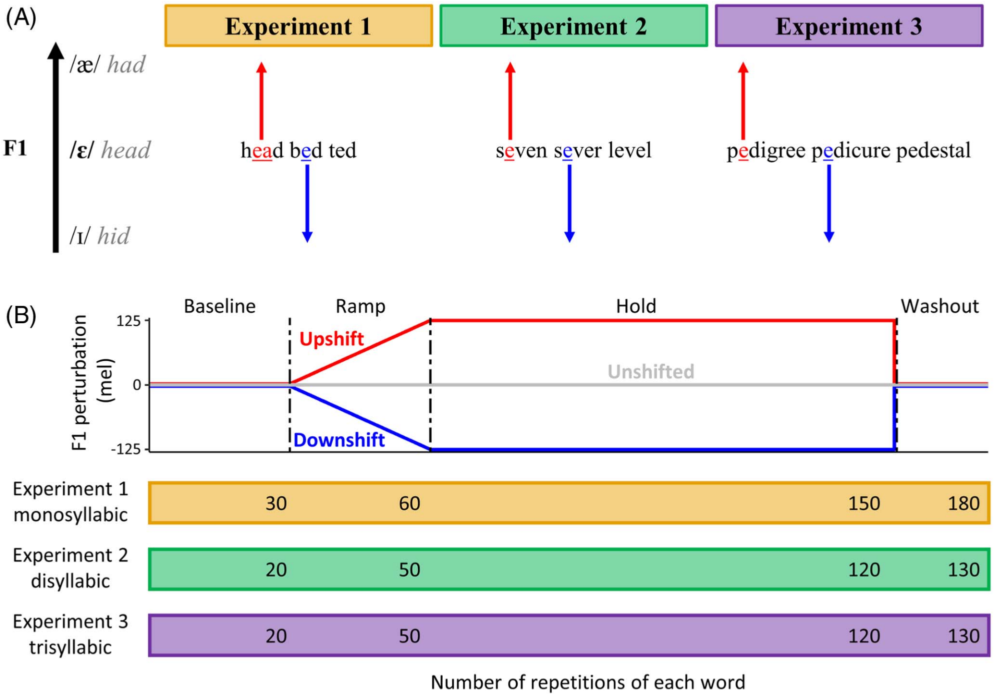
Data Processing

Vowel onset was identified as the point where periodicity was visible in the waveform and formants were visible in the spectrogram; vowel offset was identified as the point where formants, particularly F1 and F2, were no longer visible. For each vowel, F1 and F2 were tracked every 3 ms using Praat (Boersma & Weenink, 2023) via the wave_viewer package (Niziolek & Houde, 2015). Preemphasis values and linear predictive coding order were set for each participant individually; errors in formant tracking (e.g., sudden jumps, tracking wrong formants) were corrected by minimal adjustments to these values. Trials with unresolvable formant tracking errors or unintended productions (e.g., coughing, hesitations) were excluded (Experiment 1: $M = 1.22\%$, $SD = 1.46\%$, range = [0%, 3.52%]; Experiment 2: $M = 0.83\%$, $SD = 1.05\%$, range = [0%, 3.33%]; Experiment 3: $M = 0.44\%$, $SD = 0.57\%$, range = [0%, 1.92%]).

The F1 change in response to F1 perturbation is the primary dependent variable in the present study. For most stimuli, measurements from 25% to 75% into a vowel were averaged and converted

Figure 2

Experimental Design and Procedure



Note. (A) Examples of F1 perturbation assignment to words in each experiment. (B) Timeline of each experiment. F1 = first formant. See the online article for the color version of this figure.

to the mel scale to obtain a single formant value for each trial. The exception was the word *level*: In this case, the identified vocalic portion of the speech signal included both the target vowel and the initial consonant /l/ due to the difficulty in consistently segmenting these two sounds. To account for this, we used 40%–75% of the vocalic chunk to collect mean formant values. This method included the steady-state portion of the vowel and excluded the initial /l/. To capture the F1 change for each speaker (ΔF1), the average F1 values were normalized relative to the mean F1 in the last 10 trials of the baseline phase for each word.

We also measured fundamental frequency (the rate at which vocal folds vibrate; f_0) and intensity from trials with F1 data. Outliers, either 3 *SD*s away from the mean or as an abrupt sample within a trial, were removed. Next, a single average value was calculated from 25% to 75% into a vowel (the same window used for F1) for f_0 and intensity, respectively. To capture the change for each speaker, the average values were normalized as percentage change relative to the last 10 trials of the baseline phase for each word.

Statistical Analysis

Statistical analysis used mixed-effects models via the lme4 package (Bates et al., 2023; Version 1.1-34) in R (R Core Team, 2023; Version 4.3.1). Reported *p* values were calculated by the lmerTest package (Kuznetsova et al., 2020; Version 3.1-3). The α level was set at .05 in all analyses. When there were more than two independent variables, the best model was selected by running the buildmer package (Voeten, 2023; Version 2.9), which systematically compares models that differ in only one term and orders these terms according to their contribution to the overall fit (Matuschek et al., 2017). This ordering of terms achieved two goals: (a) The most important terms are the first to be included in forward fitting, and (b) the least important terms are the first to be eliminated in backward fitting. The ordering of terms and model selection used the Bayesian information criterion, which balances fit and complexity. To avoid inflating the Type I error rate (Barr et al., 2013), the initial formula that fed the model selection procedure included all possible interactions of the independent variables as fixed effects; the random effects included both participant and grouping variables (such as word) as random intercepts, plus a full random-slope structure. The initial formulae (see below) may not be justified by the experimental design, but the model selection procedure would reach a solution supported by the design and the data while also considering each term's contribution to model fit. Specifically, the model selection procedure first performed forward fitting toward a specified initial formula until a maximal possible model (without convergence failure or singular fit) was reached. After finding a maximal possible model, a backward-fitting procedure then eliminated terms that did not contribute to model fit. Unless otherwise specified, discrete independent variables were sum coded so the intercept in a selected model corresponded to the grand average across all conditions. The phia package (de Rosario-Martinez et al., 2015; Version 0.2-1) was used for post hoc comparisons; the reported *p* values used the correction procedure in Holm's (1979) study.

Because the proper way to estimate effect size for mixed-effects models is still under development (Correll et al., 2022), we report R^2 (Nakagawa et al., 2017), representing the total variance explained by the selected model and the individual terms. Cohen's *d* is reported as a standardized effect size measure for the difference between two group means. Summary statistics report marginal mean values from statistical modeling and standard errors.

The primary analysis for each experiment evaluates the presence of differential adaptation (the separation of the downshift and upshift conditions), critical for distinguishing Hypothesis 1 (syllable planning only) and Hypothesis 2 (suprasyllabic planning). This analysis examines the trial-wise F1 change in production (ΔF1) toward the end of learning (the last 10 blocks of the hold phase) and the maintenance of this learning immediately after removing the perturbation (the first 10 blocks of the washout phase; Figure 2B). Such aftereffects are a hallmark of sensorimotor adaptation as they unambiguously indicate learned changes to feedforward motor programs rather than an online reflexive response to the perturbation. Specifically, for Experiments 2 and 3, the trial-wise ΔF1 data from the last 10 blocks in the hold phase and the first 10 blocks in the washout phase were analyzed, but because Experiment 1 had more hold blocks than Experiments 2 and 3, adaptation was measured from Blocks 61 to 70, such that the number of hold blocks was identical across experiments (see the shaded areas in result figures). The initial formula had the following form, with three levels for the independent variable direction (upshift, downshift, and unshifted) and two levels for phase (hold and washout).[1] Differential adaptation should surface as a significant difference between the upshift and downshift conditions. Separate *t* tests were used to determine whether each direction differed significantly from zero:

$$\begin{aligned}\Delta\mathrm{F1} \sim \mathrm{Direction} * \mathrm{Phase} &+ (\mathrm{Direction} * \mathrm{Phase}|\mathrm{Speaker})\\ &+ (\mathrm{Direction} * \mathrm{Phase}|\mathrm{Word}).\end{aligned} \tag{1}$$

Additionally, as a supplementary analysis, generalized additive mixed modeling (GAMM), using the mgcv (Wood, 2023; Version 1.9-0) and itsadug (van Rij et al., 2022; Version 2.4.1) packages, analyzed the ΔF1 of all trials in each experiment to evaluate when the upshift and downshift conditions diverged. Because fitting maximally complex models is inappropriate for GAMM (Baayen et al., 2017), we followed the model-building procedure described by Wieling (2018). The Supplemental Materials contain the details of the GAMM modeling.

We also examined whether the word items could predict the adaptation size within each experiment given that (a) a word's acoustic realization is known to be affected by its lexical frequency (e.g., Pluymaekers et al., 2005), and the words in our experiments differ in their lexical frequency and (b) more frequent words are often hypothesized to form a single "chunk" (e.g., Guenther, 2016), which may influence the adaptation responses. This analysis relates specifically to perturbed words, so only the upshift and downshift

---

[1] Formulae of statistical models in this article follow R syntax, where the variable to the left of the tilde (~) is the dependent variable and the independent variables are to the right. An interaction of two independent variables is joined by a colon (e.g., Direction:Phase). When two independent variables are joined by an asterisk (*), its full expression includes individual variables and all their interaction (e.g., Direction * Phase = Direction + Phase + Direction:Phase). For mixed-effects models, fixed effects do not use parentheses, and random effects are placed within parentheses. Within the parentheses, a random intercept is placed to the right of a pipe (|), and its corresponding random slope is specified to the left.

conditions were included. Because F1 is expected to decrease in the upshift (negative F1 change) but increase in the downshift (positive F1 change), we sign-flipped (multiplied by -1) the F1 change values in the upshift such that the expected values were always positive (labeled $\Delta F1_{\text{word}}$ in the formula below). The potential effect of lexical frequency[2] was examined by placing word as a fixed effect; all lexical frequency estimates were from Google Ngram Viewer (Orwant & Brockman, 2019), using its most recent reports in 2019. We note that this analysis is exploratory as lexical frequency is only one component that differs between the two stimuli in each experiment, the stimuli were not designed with lexical frequency differences in mind, and even if there is a lexical frequency effect on the adaptation behavior, the adaptation results of our primary analyses should not be affected by lexical frequency, as our design employed counterbalancing of the stimuli. The initial formula fed to model selection had the following form:

$$\Delta F1_{\text{word}} \sim \text{Word} * \text{Phase} * \text{Direction} + (\text{Word} * \text{Phase} * \text{Direction}|\text{Speaker}). \quad (2)$$

To preview, we found differential adaptation in all three experiments. Notably, Experiment 3 in particular is a strong test of suprasyllabic motor plans, as the trisyllabic words in Experiment 3 had phonemically identical second syllables ("di"), minimizing the likelihood that acoustic or articulatory differences in the first syllables caused by coarticulation may provide distinct contexts that may enable differential adaptation. We further examined the baseline F1, F2, and F3 of the first syllables in Experiment 3 to verify the absence of such accidental contextual differences (the last 10 repetitions of the baseline phase; formulae below):

$$F1|F2|F3 \sim \text{Word} + (\text{Word}|\text{Speaker}). \quad (3)$$

F3 measures were obtained from Audapter using the same analysis window as the F1 and F2; the median F3 of each trial was extracted to minimize the influence of outliers.

In addition to establishing whether differential adaptation occurred in response to opposing perturbations in each experiment (Hypothesis 1 vs. Hypothesis 2), a critical comparison concerns the size of the differential adaptation (downshift–upshift) across experiments; a difference in the size of differential adaptation across experiments would suggest an interaction between syllable and word planning (Hypothesis 3). To compare the same measure across studies, we used the difference of F1 change between the upshift and downshift conditions (downshift–upshift; labeled $\Delta F1_{\text{diff}}$ henceforth). For each participant, a single measure of differential adaptation was calculated in the analyzed hold and washout windows as described above, where the amount of perturbation exposure during the hold phase was identical across experiments. This analysis started with the initial formula below; the variable experiment was coded as an ordered factor to reflect the number of syllables in the stimuli in each experiment:

$$\Delta F1_{\text{diff}} \sim \text{Phase} * \text{Experiment} + (\text{Phase} * \text{Experiment}|\text{Speaker}). \quad (4)$$

To preview, the above analysis revealed a larger size of differential adaptation in Experiment 1 than in the other two experiments (Experiment 1 > Experiments 2 and 3, i.e., monosyllabic > multisyllabic). However, we also expect longer vowels in the monosyllabic words than in the multisyllabic words (Umeda, 1975), which was confirmed in a separate analysis (duration: Experiment 1 > Experiment 2 > Experiment 3, all $p < .001$; see Supplemental Material for details). Studies of sensorimotor adaptation in reaching report that learning is greater with continuous visual feedback of hand position compared with more limited feedback about the reach endpoint (e.g., Schween et al., 2014). While this may suggest that longer exposure to a perturbation could lead to greater adaptation, (a) to our knowledge, there is not any direct comparison of shorter versus longer movements with continuous feedback in the reaching literature (analogous to the differences in vowel duration), and (b) advancing the endpoint feedback in time to occur during the reach while maintaining the (short) duration of that feedback results in equivalent adaptation to continuous feedback (Wang et al., 2024), suggesting that the timing of feedback during movement is more critical for learning than longer exposure duration per se. Nonetheless, we tested whether duration could play a modulatory role in the size of differential adaptation. We first manually added the mean duration of the perturbed words per speaker per phase ($\text{Duration}_{\text{mean}}$) in the selected model, but model diagnostics indicated that this formula was misspecified due to multicollinearity (variance inflation factor > 5), driven by the strong correlation of experiment and $\text{Duration}_{\text{mean}}$. We then modeled the size of differential adaptation with only $\text{Duration}_{\text{mean}}$ in the initial formula (see below) to evaluate whether it was a better predictor of $\Delta F1_{\text{diff}}$ than experiment (e.g., comparing the selected experiment-only and duration-only models directly according to Bayesian information criterion and variance explained by individual models and their constituent terms):

$$\Delta F1_{\text{diff}} \sim \text{Phase} * \text{Duration}_{\text{mean}} + (\text{Phase} * \text{Duration}_{\text{mean}}|\text{Speaker}). \quad (5)$$

Last, Experiments 2 and 3 showed a global tendency to increase F1 across stimuli, regardless of the perturbation direction. Even in Experiment 1, the unshifted condition has a positive F1 change, a tendency also seen in previous work using the same paradigm (Rochet-Capellan & Ostry, 2011). Global increases in F1 often accompany clearly articulated speech (clear speech, henceforth), as produced, for example, in the presence of background noise or when speaking to a listener with compromised hearing abilities (e.g., hearing loss, or from a different language background; Lindblom, 1990). Because clear speech often occurs when there is communication difficulty, the masking noise and formant perturbations in our experiments may have induced similar behavior. To test this possibility, we additionally report changes in other speech parameters often associated with clear speech, including f_0, intensity, and duration (Krause & Braida, 2004).

Transparency and Openness

This study follows APA Style Journal Article Reporting Standards (Appelbaum et al., 2018). We report how we determined our sample

---

[2] In the single-experiment analysis described above, the frequency analysis described here, and the cross-experiment analysis described below, we also included log-transformed lexical frequency estimates in the initial formula. The results regarding differential adaptation were qualitatively identical to the results we reported here (see Supplemental Materials for details).

size, all data exclusions (if any), all manipulations, and all measures in the study. All data, analysis code, and research materials are publicly available on the Open Science Framework at https://osf.io/ytwqp/. All experiments and analyses can be reconstructed using the code shared at the same link; this link also contains the Supplemental Materials. This study's design and its analysis were not preregistered. The University of Wisconsin–Madison's Institutional Review Board approved all procedures.

Results

We report results from three experiments testing the scope of speech motor planning via adaptation to altered auditory feedback. In Experiment 1, opposing perturbations are applied to the same vowel in different monosyllabic words. In Experiments 2 and 3, opposing perturbations are applied to the same initial syllable in different multisyllabic words. If the word is a motor planning unit independent of the syllable, differential adaptation should occur in all three experiments to an equal degree (Hypothesis 1). Conversely, if speech motor planning relies mainly on syllables and the word is not a motor planning unit, speakers will exhibit differential adaptation in Experiment 1, but not in Experiments 2 and 3, where the learning from opposing perturbations to the same syllable would cancel out (Hypothesis 2). Last, if the word is a motor planning unit that interacts with the syllable, we expect differential adaptation in all experiments but of a reduced size in Experiments 2 and 3 relative to Experiment 1, as syllable-level learning (no adaptation) conflicts with word-level differential adaptation in these two experiments (Hypothesis 3).

Experiment 1

As expected, participants adapted differentially to the opposing perturbations applied to distinct monosyllabic words (Rochet-Capellan & Ostry, 2011). When F1 was perturbed upward, speakers lowered their produced F1; when F1 was perturbed downward, speakers raised their produced F1. Adaptation reached a plateau toward the end of the hold phase and then declined after removing the perturbation during the washout phase (Figure 3A). According to the GAMM analysis, the downshift and upshift separation started in Block 49 (during the Ramp phase); these two conditions remained separated for the remainder of the experiment. In the blocks analyzed (shaded areas in Figure 3A), all individual speakers' responses mirrored the group average (Figure 3B): 100% produced a higher mean F1 in the downshift condition compared with the upshift condition in the hold phase, and 93% maintained this pattern in the washout phase. The selected model ($R^2 = 0.216$) for the analysis of ΔF1 was the following:[3]

$$\Delta \mathrm{F1} \sim \mathrm{Direction} + \mathrm{Phase} + (\mathrm{Direction} + \mathrm{Phase}|\mathrm{Speaker}). \quad (6)$$

Model results indicated that the F1 change was significantly different between the perturbation conditions, a main effect of direction, $\chi^2(2) = 44.97$, $R^2 = 0.343$, $p < .001$. All three perturbation conditions differed significantly from each other (all $p < .001$; upshift vs. downshift: $d = 1.31$). Adaptation differed from zero in the upshift (−22.79 ± 2.34 mels, $p = .005$) and the downshift (33.23 ± 2.63 mels, $p < .001$) conditions, but not in the unshifted (5.06 ± 2.45 mels, $p = .55$). Phase was not significant in the selected model, $\chi^2(1) = 0.001$, $R^2 = 0.001$, $p = .97$, suggesting a limited change from the hold to the washout phases.

We also examined whether lexical items influenced the adaptation size in Experiment 1. The selected model had the following formula ($R^2 = 0.022$), which did not contain the variable "word," suggesting individual words did not differ in adaptation:

$$\Delta \mathrm{F1}_{\mathrm{word}} \sim \mathrm{Phase} + \mathrm{Direction} + (\mathrm{Direction}|\mathrm{Speaker}). \quad (7)$$

In sum, Experiment 1 replicates the principal findings of Rochet-Capellan and Ostry (2011): Speakers adapt separately to opposing perturbations in monosyllabic words containing the same vowel, and this separation was maintained during the washout phase. Importantly, we found this behavior even when adjacent trials did not have identical perturbation/word; the ability to adapt to the opposing perturbations, in this case, cannot be attributed to sequential exposure to the same perturbation, a feature that could drive differential adaptation even in the absence of planning cues (Osu et al., 2004).

Experiment 2

Experiment 2 tested whether participants could adapt to opposing perturbations applied to the same initial syllable "sev" in different disyllabic words (*seven* and *sever*). Speakers in Experiment 2 responded differently to the upshift and downshift perturbations. However, they tended to increase F1 regardless of the perturbation direction. Still, the F1 increase in the downshift was greater than in the upshift. Intriguingly, in Experiment 2, while speakers' F1 production in the upshift condition initially increased in parallel with the downshift condition, the mean F1 change switched to the predicted direction (from positive to negative) toward the end of hold and during washout (Figure 4A). According to the GAMM analysis, the downshift and upshift conditions first separated in Block 87 (during the hold phase) and remained separated for the remainder of the experiment. Compared with the early separation of the upshift and downshift in Experiment 1, the differential adaptation in Experiment 2 took longer to emerge, potentially due to a conflict between syllable-level learning (no adaptation) and word-level learning (differential adaptation). In the blocks analyzed (shaded areas in Figure 4A), speakers were largely consistent in their response to the auditory perturbations (Figure 4B): 95% produced a higher mean F1 in the downshift condition compared with the upshift condition in the hold phase and 85% in the washout phase. The final selected model ($R^2 = 0.024$) was:

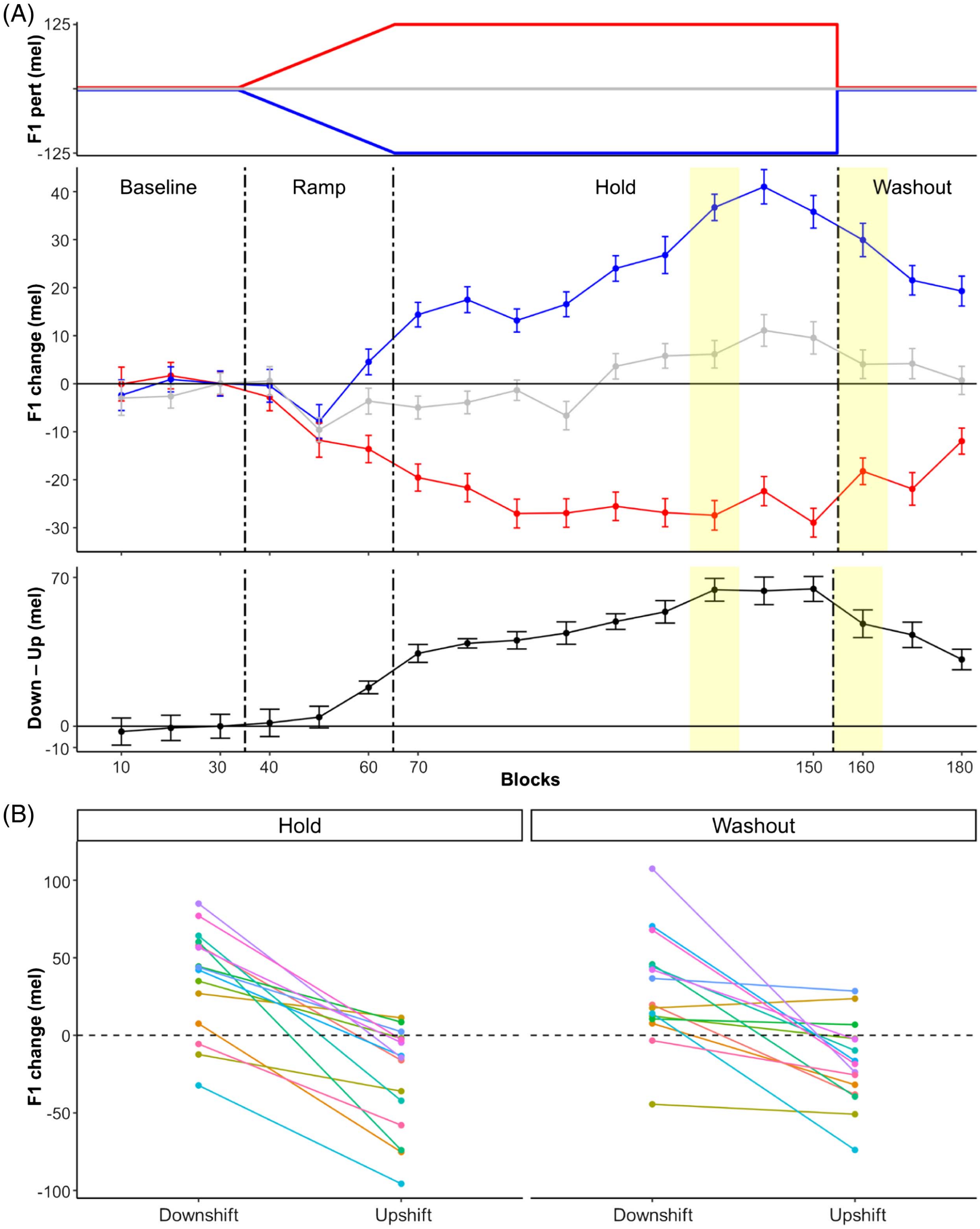
$$\Delta \mathrm{F1} \sim \mathrm{Direction} + \mathrm{Phase} + (\mathrm{Direction} + \mathrm{Phase}|\mathrm{Speaker}). \quad (8)$$

Speakers adapted according to the perturbation they received, a main effect of direction, $\chi^2(2) = 16.06$, $R^2 = 0.033$, $p < .001$. The upshift differed significantly from the downshift ($d = 0.39$, $p < .001$), but the other two differences were not significant (downshift vs. unshifted: $d = 0.30$, $p = .20$; upshift vs. unshifted: $d = 0.07$, $p = .62$). When compared against zero, none of the three perturbation conditions differed significantly from zero (upshift = −4.34 ± 2.29 mels, $p = .99$; downshift = 13.08 ± 2.33 mels, $p = .34$; unshifted = −1.22 ± 2.55 mels, $p = .99$). Phase was not significant in the selected

---

[3] The independent variables of the selected models reported in this article are listed according to their contribution to the overall model fit in descending order, measured in the significance of log-likelihood change.

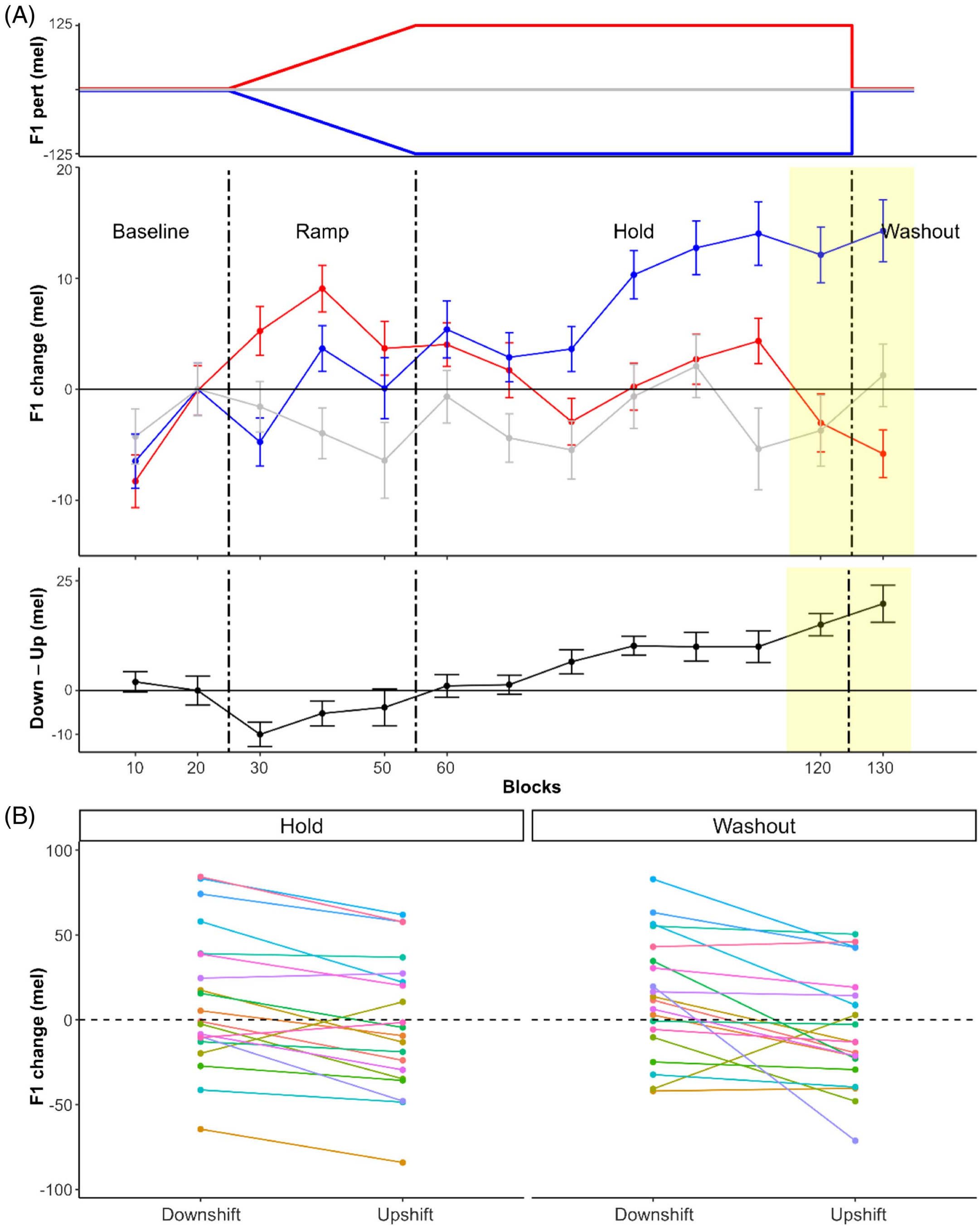
Figure 3
Experiment 1 Results



Note. Individual data points show means and standard errors across 10 blocks. (A) the F1 perturbation, F1 change, and differential adaptation (downshift–upshift). Colors show different perturbation conditions: red shows the upshift, blue shows the downshift, and gray shows the unshifted. Trials in the shaded areas were included in the primary statistical analysis; Blocks 61–70 were selected to achieve an identical amount of perturbation exposure in the hold phase across experiments. (B) Individual speakers' F1 change in the analyzed hold and washout windows. Colors represent individual speakers. F1 = first formant. See the online article for the color version of this figure.

Figure 4

Experiment 2 Results



Note. Individual data points show means and standard errors across 10 blocks. (A) The F1 perturbation, F1 change, and differential adaptation. Colors show different shift conditions: red shows upshift, blue shows downshift, and gray shows unshifted. Trials in the shaded areas were included in the primary statistical analysis. (B) Individual speakers' F1 change in the analyzed hold and washout windows. Colors represent individual speakers. F1 = first formant. See the online article for the color version of this figure.

model, $\chi^2(1) = 0.124$, $R^2 < 0.001$, $p = .72$, suggesting little change between the hold and washout phases.

Similar to Experiment 1, we examined whether lexical items influence adaptation size in Experiment 2. The selected model had the following formula ($R^2 = 0.002$), where word was the only predictor in the selected model but was not significant, $\chi^2(1) = 0.08$, $R^2 = 0.002$, $d = 0.09$, $p = .78$:

$$\Delta F1_{\text{word}} \sim \text{Word} + (1 + \text{Word}|\text{Speaker}). \tag{9}$$

In sum, we tested differential adaptation to the identical initial syllable of disyllabic words in Experiment 2. Participants tended to increase their F1 regardless of the formant perturbation direction initially; the production in the upshift lowered and separated from the downshift by the end of the hold phase, and this separation was maintained during the washout phase. Overall, speakers do learn to adapt differentially to the two perturbations, but this separation emerged later than in the monosyllabic words in Experiment 1.

Experiment 3

While the differential adaptation found in Experiment 2 suggests word-level motor plans, one might suspect that the distinct second syllables in Experiment 2 (“ver” vs. “ven”) may accidentally introduce baseline first-syllable differences due to coarticulation; in such a scenario, distinct movement kinematics in the first syllable, rather than the distinct disyllabic contexts, could have enabled differential adaptation. Indeed, post hoc analysis found that the first syllable of *seven* had a significantly higher baseline F2 than its counterpart in *sever* (the last 10 repetitions of the baseline phase; $p < .001$); there were no differences in either F1 or F3 ($p > .73$). Experiment 3 was designed to better isolate movement planning from kinematic contexts by examining the differential adaptation of the first syllable in trisyllabic words with phonemically identical second syllables, thus minimizing the potential influence of kinematic differences in the (shared) first syllable. Specifically, Experiment 3 tested whether participants could adapt to opposing perturbations applied to the same initial syllable “ped” in different trisyllabic words (*pedigree* and *pedicure*). We first confirmed that our stimuli had the intended effect of minimizing coarticulatory movements in the first syllables. Indeed, baseline F1, F2, and F3 did not differ between words: The predictor word was absent in all selected formulae. Therefore, differential adaptation in Experiment 3, if observed, is unlikely to have been driven by coarticulation from the later syllables to the first syllables.

Like Experiment 2, speakers in Experiment 3 tended to increase F1 initially regardless of the perturbation direction; despite this global trend, the F1 increase in the downshift was greater than in the upshift, thus showing differential adaptation (Figure 5A). According to the GAMM analysis, the downshift and upshift first separated in Block 63 (during the hold phase) and remained separated for the remainder of the experiment. In the analyzed blocks (shaded areas in Figure 5A), most speakers mirrored the group average (Figure 5B): 95% produced a higher mean F1 in the downshift condition compared with the upshift condition in the hold phase and 90% in the washout phase. The selected model ($R^2 = 0.040$) for analyzing the F1 change was the following:

$$\Delta F1 \sim \text{Direction} + \text{Phase} + (\text{Direction}|\text{Speaker}). \tag{10}$$

Speakers adapted according to the perturbation they received, a main effect of direction, $\chi^2(2) = 13.37$, $R^2 = 0.05$, $p = .001$. The downshift differed significantly from the upshift ($d = 0.43$, $p = .001$) and the unshifted ($d = 0.38$, $p = .004$), but the difference between the upshift and the unshifted was not significant ($d = 0.09$, $p = .29$). When compared against zero, only the downshift differed significantly from zero (upshift = 0.92 ± 1.82 mels, $p = .87$; downshift = 15.55 ± 1.57 mels, $p = .010$; unshifted = 4.04 ± 1.45 mels, $p = .66$). Consistent with the impression that F1 tended to increase throughout the experiment, the overall F1 change across all three conditions was also larger in the hold phase than in the washout phase, a main effect of phase, $\chi^2(1) = 10.58$, $R^2 = 0.005$, $d = 0.14$, $p = .001$, but in neither phase did the overall F1 differ significantly from zero (hold = 9.06 ± 1.27 mels, $p = .11$; washout = 4.58 ± 1.41 mels, $p = .33$).

Again, we examined whether lexical items influence adaptation size in Experiment 3. The selected model ($R^2 = 0.030$) was:

$$\Delta F1_{\text{word}} \sim \text{Word} + \text{Phase} + \text{Word}:\text{Phase}(\text{Word}|\text{Speaker}). \tag{11}$$

Word was not significant in the model, $\chi^2(1) = 1.03$, $R^2 = 0.024$, $d = 0.31$, $p = .31$. Although the interaction of word and phase was significant, $\chi^2(1) = 8.27$, $R^2 = 0.005$, $p = .004$, this was driven by *pedicure* having larger adaptation during washout than during hold ($d = 0.21$, $p = .003$) rather than by the difference between the words in each phase (both $p > .28$).

As an exploratory analysis, we also examined the F1 change in the second syllables of the perturbed words in Experiment 3 using the same model structure and analysis pipeline as for the first syllable. The selected model was the following ($R^2 = 0.026$):

$$\Delta F1 \sim \text{Direction} + \text{Phase} + (\text{Direction} + \text{Phase}|\text{Speaker}). \tag{12}$$

The downshift condition had significantly greater F1 change than the upshift condition, a significant main effect of direction, $\chi^2(1) = 5.22$, $R^2 = 0.025$, $d = 0.033$, $p = .022$. When compared against zero, neither direction differed significantly from zero (upshift = −4.32 ± 1.87 mels, $p = .54$; downshift = 8.02 ± 1.92 mels, $p = .51$). No other term in the selected model was significant ($p > .19$). Although the direction of the downshift–upshift separation is the same as in the first syllables, the source of this change is unclear: While the second-syllable separation could be caused by the adaptation of the motor plan for the second syllable, it may also result, more simply, from carryover coarticulatory effects of the change in the first syllable, or from a combination of these two mechanisms.

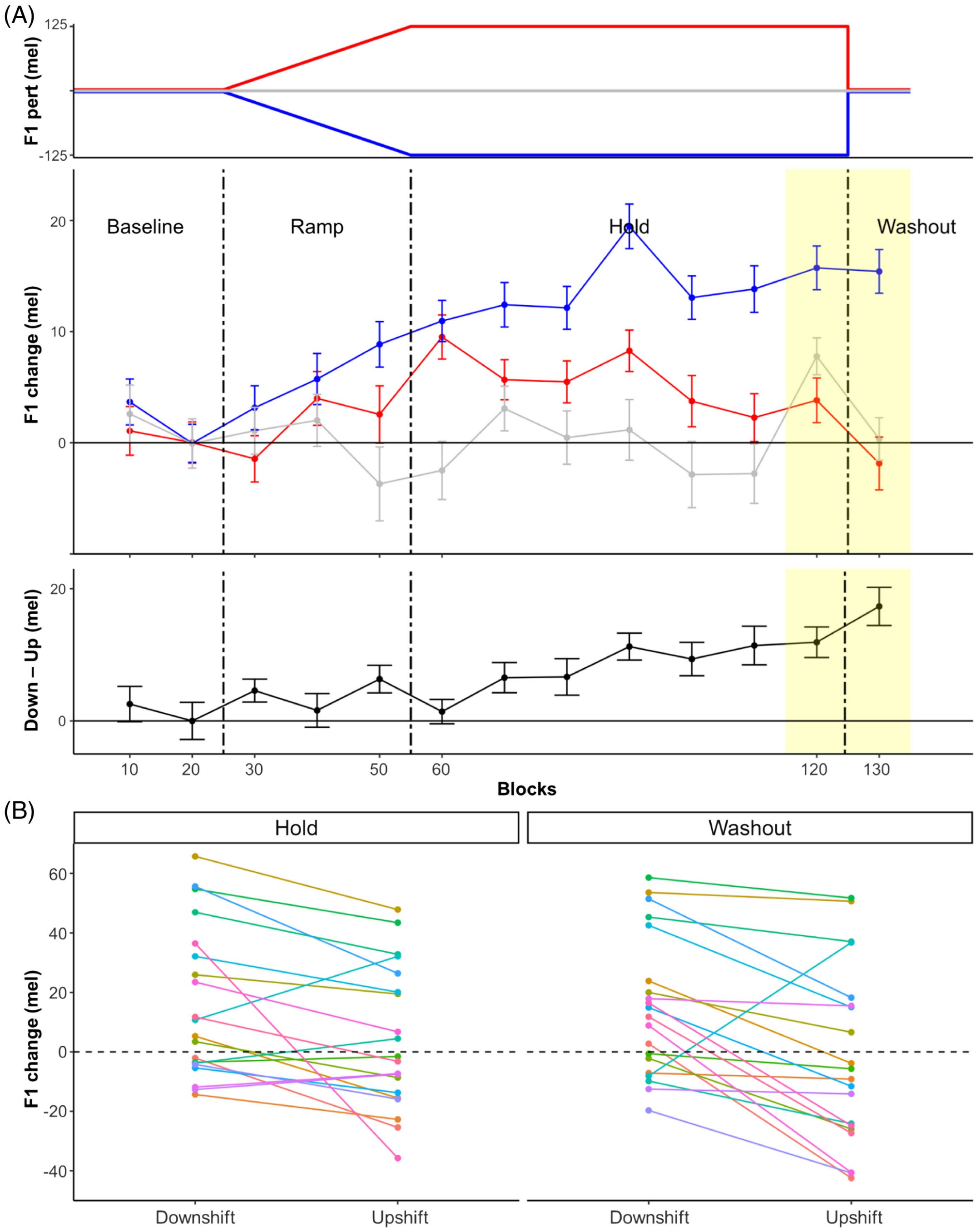
To summarize, we tested adaptation to opposing perturbations in trisyllabic words in Experiment 3. Participants tended to increase their F1 regardless of the formant perturbation direction initially; the production in the upshift then lowered and separated from the downshift. Overall, speakers do learn to adapt differentially to the two perturbations. Crucially, the differential adaptation in Experiment 3 cannot be attributed to coarticulatory-driven kinematic differences in the first syllables.

Comparisons of Differential Adaptation Across Experiments

Comparing the three experiments, there is a stark distinction between the size of adaptation in the monosyllabic words in Experiment 1 and the multisyllabic words in Experiments 2 and 3

Figure 5

Experiment 3 Results

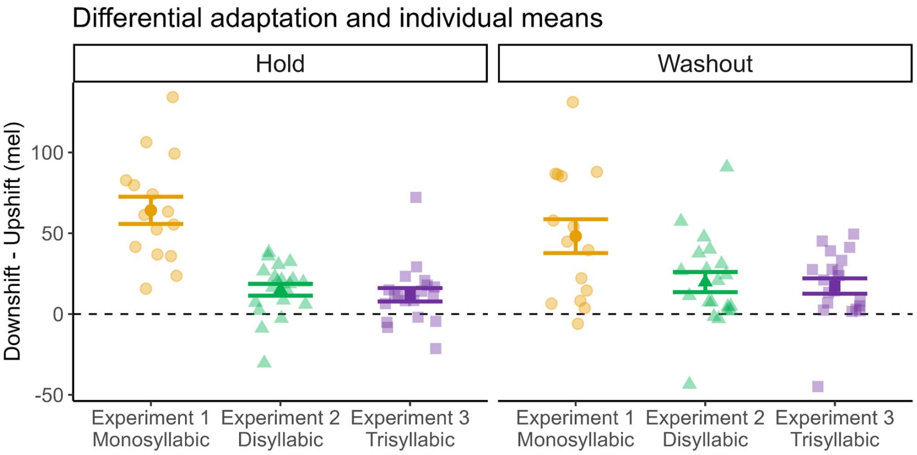


Note. Individual data points show means and standard errors across 10 blocks. (A) The F1 perturbation, F1 change, and differential adaptation. Colors show different shift conditions: red shows upshift, blue shows downshift, and gray shows unshifted. Trials in the shaded areas were included in the primary statistical analysis. (B) Individual speakers' F1 change in the analyzed hold and washout windows. Colors represent individual speakers. F1 = first formant. See the online article for the color version of this figure.

(Figure 6). In the monosyllables in Experiment 1, differential adaptation (downshift–upshift) plateaued at approximately 60 mels toward the end of the hold phase and decreased to about 45 mels after removing the perturbation during the washout (Figure 3A). By contrast, in the multisyllabic words in Experiments 2 and 3, differential adaptation reached a maximum of approximately 20 mels at the end of the hold phase without decreasing during the washout phase (Figures 4A and 5A). The variance explained by the selected

Figure 6

Differential Adaptation (Downshift–Upshift) Across Experiments in the Analyzed Hold and Washout Windows



Note. The transparent dots show individual participant means. Error bars show standard errors. See the online article for the color version of this figure.

model in each experiment was also markedly greater in Experiment 1 than in Experiments 2 and 3, suggesting more variable responses in the multisyllabic words. Comparing the size of differential adaptation across experiments yielded the following selected model ($R^2 = 0.418$):

$$\Delta F1_{\text{diff}} \sim \text{Experiment} + \text{Phase} + \text{Experiment}:\text{Phase} + (1|\text{Speaker}). \tag{13}$$

Unsurprisingly, the size of differential adaptation differed across experiments, a significant main effect of experiment, $\chi^2(2) = 63.58$, $R^2 = 0.477$, $p < .001$. Experiment 1 (62.17 ± 6.77 mels) had a larger differential adaptation than Experiment 2 (17.39 ± 3.58 mels; $d = 1.31$, $p < .001$) and Experiment 3 (8.44 ± 3.15 mels; $d = 1.46$, $p < .001$), but the difference between Experiments 2 and 3 was not significant ($d = 0.13$, $p = .24$). The model also contained a significant interaction of experiment and phase, $\chi^2(2) = 9.86$, $R^2 = 0.032$, $p = .007$. Post hoc comparisons found that this interaction was primarily driven by a significant decline of differential adaptation in Experiment 1 (hold > washout, $p = .016$), which was absent in Experiments 2 and 3 (both $p > .54$). No other term in the model was significant ($p = .52$).

These results show a larger differential adaptation in the monosyllabic words than in the multisyllabic words, with no difference between the multisyllabic words. While there is little direct evidence that the duration of exposure to altered feedback within a single movement/trial affects learning, it is nonetheless conceivable that the difference in the differential adaptation size could be driven not by the mono- versus multisyllabic distinction but by longer vowel durations in the monosyllabic words (Umeda, 1975). As expected, we found this pattern in our data (vowel duration: Experiment 1 > Experiment 2 > Experiment 3; Experiment 1 = 247.36 ± 13.66 ms, Experiment 2 = 121.79 ± 4.45 ms, Experiment 3 = 61.89 ± 2.26 ms; $p < .001$ in all pairwise comparisons). Because vowel duration and experiment were highly collinear, we could not include both variables in a single model (variance inflation factor > 5). Accordingly, we checked if duration was a better predictor of the differential adaptation size than experiment by running a separate model selection procedure with only $\text{Duration}_{\text{mean}}$ in the initial formula, which resulted in the following model ($R^2 = 0.215$):

$$\Delta F1_{\text{diff}} \sim \text{Duration}_{\text{mean}} + \text{Phase} + \text{Duration}_{\text{mean}}:\text{Phase} + (\text{Duration}_{\text{mean}}|\text{Speaker}). \tag{14}$$

This duration-only model explained less variance than the above experiment-only model (R^2: 0.215 vs. 0.418) and had a worse fit (Bayesian information criterion: 1036.08 vs. 1010.15). As a main effect, $\text{Duration}_{\text{mean}}$, $\chi^2(1) = 9.63$, $R^2 = 0.203$, $p < .001$ was a worse predictor of $\Delta F1_{\text{diff}}$ than experiment, $\chi^2(2) = 63.58$, $R^2 = 0.477$, $p < .001$.[4] Another line of evidence also suggests that the difference in the differential adaptation size was primarily driven by the mono- versus multisyllabic distinction rather than by duration: Although Experiment 2 had a markedly longer duration than Experiment 3 ($p < .001$), to the extent that the perturbed syllables in Experiment 2 were about twice the duration in Experiment 3 (121.79 ms vs. 61.89 ms), they did not differ in the size of differential adaptation. Therefore, the smaller differential adaptation in multisyllabic words cannot be reduced to duration differences but is more likely due to multisyllabic versus monosyllabic words.

Changes in Speech Parameters Related to Clear Speech

In Experiments 2 and 3, although speakers adapted separately in the downshift and upshift conditions, there was a global tendency to increase F1. Similarly, in Experiment 1, the unshifted condition also increased F1 despite receiving no perturbation. Here, we evaluate clear speech as a potential cause of this overall F1 increase. Figure 7

---

[4] When lexical frequency was included in the initial formula, the experiment-only model was also a better predictor than the duration-only model (see the Supplemental Materials for details).

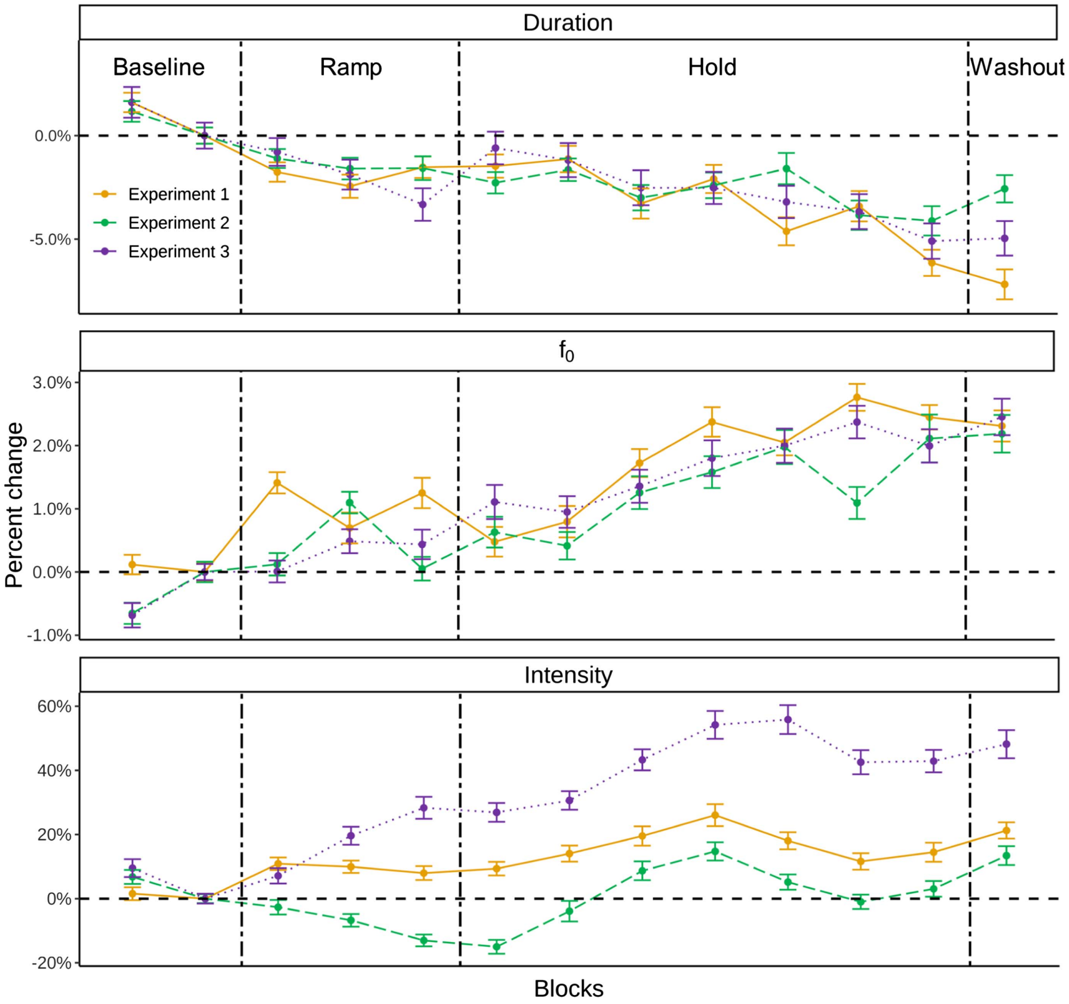
shows the percent changes in the acoustic parameters often associated with clear speech across experiments (duration, f_0, and intensity; Krause & Braida, 2004). Data for Experiment 1 were adjusted to make equal comparison across experiments possible (omitting the first 10 baseline blocks and the last 20 blocks from the hold and washout phases). The Supplemental Materials contain figures additionally showing these changes as a function of perturbation direction; generally, different perturbation directions within a single experiment were parallel.

According to Figure 7, duration had a steady decline in all three experiments. This is expected given the repeated production of a small set of stimulus words (e.g., Niziolek & Parrell, 2021), potentially nullifying any increase that would typically be associated with clear speech. By contrast, f_0 increased steadily in all three experiments, in line with a clear speech account. However, past reports have shown that f_0 tends to increase across extended repetitions of simple stimuli, like those used here (e.g., Jones & Munhall, 2000). Therefore, although the f_0 increase is consistent with a clear speech account, causes other than clear speech are also likely. The intensity changes had a greater range and more fluctuations, which may be attributed to fatigue and the inclusion of breaks in our experiment: Speakers may drop intensity due to fatigue but may resume intensity after a break. Still, the observed general increase in intensity does not contradict a clear speech account.

In brief, of the three acoustic parameters often associated with clear speech, the f_0 and intensity changes do not contradict the predictions of clear speech. The duration change is the opposite of what clear speech predicts but is expected due to the repeated production of a limited set of words. Therefore, the evidence is mixed regarding whether the global F1 increase in Experiments 2 and 3, as well as the F1 increase in the unperturbed stimuli here and in Rochet-Capellan and Ostry (2011), is attributable to clear speech.

Figure 7

Change in Duration, Fundamental Frequency (f_0), and Intensity in All Experiments



Note. Individual data points show means and standard errors across 10 blocks. Because Experiment 1 had more trials than Experiments 2 and 3, data for Experiment 1 exclude the first 10 baseline blocks, the final 20 hold blocks, and the final 20 washout blocks. See the online article for the color version of this figure.

Perturbation Awareness

After the main experiment, participants filled out a questionnaire, where they were first asked to recall and describe the task they did. They were then informed that there were two groups, one receiving true auditory feedback and the other receiving manipulated auditory feedback. It was also highlighted that the manipulation was made to their voice. Participants were asked to choose which group they thought they were in. Most participants believed they received true auditory feedback (53.33% in Experiment 1, 75% in Experiment 2, and 70% in Experiment 3). Of the participants who reported they were in the manipulated group, some identified a change to their vowel characteristics (50% in Experiment 1, 25% in Experiment 2, and 50% in Experiment 3). There is strong evidence that speech sensorimotor adaptation is a largely implicit process and that even when participants are aware of the perturbation, they are unable to generate any strategies to oppose it (Kim et al., 2020; Munhall et al., 2009). Therefore, participants developing conscious strategies is an unlikely cause for the differential adaptation in our study.

Discussion

Differential Adaptation Supports Suprasyllabic Motor Plans

In three experiments, we tested whether speakers could adapt differentially to opposing perturbations applied to words sharing potential planning units (Experiment 1: monosyllabic words sharing the same vowel; Experiments 2 and 3: multisyllabic words sharing the first syllable). All three experiments showed a separation between the upshift and downshift conditions during the end of the hold phase; these separations were also maintained after removing the perturbation during the washout phase (all $p \leq 0.001$), strengthening the idea that the separation between the downshift and the upshift was not merely the result of online reflexive responses but differential adaptation due to updated motor plans.

Importantly, the differential adaptation in the multisyllabic words, especially in Experiment 3, cannot be attributed purely to kinematic differences in the first syllables that may inadvertently provide distinct contexts to enable differential adaptation. Our design ensured that carryover coarticulation was absent (identical initial syllables in the multisyllabic words) and differences in anticipatory coarticulation were minimized (the same segment following the initial syllables). Particularly in Experiment 3, the two target words differed only in the third syllable: Because anticipatory coarticulation is largely restricted to adjacent syllables (Recasens, 2018), we expect essentially no coarticulation in the first syllable in this case. We further confirmed that coarticulation in the target vowels was absent in the first syllables of Experiment 3: Baseline formant values (F1, F2, and F3) did not differ between the target stimuli. Therefore, the separation of the upshift and the downshift conditions in Experiment 3 must be attributed to the multisyllabic contexts rather than any preexisting, coarticulation-driven kinematic differences in the perturbed first syllables. In reaching, planning the same movement in different contexts is necessary and sufficient for differential adaptation, while distinct kinematic contexts in the absence of planning are not (Sheahan et al., 2016). By extension, a similar mechanism is likely to operate in speech in our experiments, supporting distinct suprasyllabic/word-level motor plans.

Notably, our finding shows word-level/suprasyllabic motor planning above and beyond what is shown by anticipatory coarticulation, providing more definitive evidence for suprasyllabic planning. Although anticipatory coarticulation across multiple syllables must be attributed to speech motor planning at some level (Recasens, 2018; Whalen, 1990), coarticulation itself does not provide strong evidence for suprasyllabic/word-level planning. First, the anticipatory spread of gestures is temporally limited to, at most, the vowel preceding the triggering segment, though potentially across multiple intervening consonants (e.g., Noiray et al., 2011; cf. the trisyllabic words in Experiment 3). Second, coarticulatory effects are due to either biomechanical inertia or effort optimization (Recasens, 2018). Third, coarticulation is not tied to specific words or syllable sequences but occurs consistently whenever the triggering segments are found. Together, the existing literature suggests that coarticulation reflects lower level movement optimization, potentially related to motor programming (the motoric specification of movement plans in context; Van der Merwe, 2021), instead of motor planning per se. Conversely, the differential adaptation in our experiments has no biomechanical motivation: It is unrelated to existing gestures in the following syllables, is far from the conditioning syllable (in Experiment 3), and is tightly tied to a specific segmental/syllabic context within a word. Therefore, the differential adaptation in our experiments provides more direct, unambiguous support for word-level/suprasyllabic motor planning than coarticulation.

Although our results indicate that multisyllabic word production involves some level of suprasyllabic planning, given the stimuli used in our experiments, it is unclear whether this planning context is provided by words per se or simply by the repeated production of the same multisyllabic sequence. While it is possible that our results would hold for any multisyllabic sequence (e.g., metrical foot, multisyllabic word, multiword phrase), both the psycholinguistic and speech motor planning literature suggest that words (or, at least frequent words) play a critical role in planning (e.g., Levelt, 2001; Tourville & Guenther, 2011); thus, we would predict that the current results would hold only for words and not for multiword sequences. Further experiments are needed to clarify the precise nature of the suprasyllabic context that enables the differential adaptation as observed in the present study.

Reduced Differential Adaptation in Multisyllabic Words Suggests Syllable-Word Interaction During Planning

The differential adaptation in the multisyllabic words (disyllables: $d = 0.39$; trisyllables: $d = 0.43$) was significantly smaller than in the monosyllabic words ($d = 1.31$). The reduced differential adaptation in the multisyllabic words is explainable if planning occurs at both the suprasyllabic (possibly word) and the syllable levels, where syllabic-level plans are recombined across word context. Each word forms an independent syllable in the monosyllabic words, so there is no conflict between word- and syllable-level planning. In the multisyllabic words, the perturbed syllable is identical; the conflict between syllable-level learning (opposing adaptive movements cancel out) and word-level learning (enabling differential adaptation) could explain the reduced differential adaptation size. This cross-level conflict could also explain the later separation of the upshift and downshift conditions in the multisyllabic words relative to the monosyllabic words.

Strikingly, our results differ from previous results in upper limb control, where the adaptation of a single movement in multimovement sequences is roughly equivalent to the adaptation of the same movement on its own (e.g., Howard et al., 2015; Sheahan et al., 2016). Conflicts between syllable- and word-level planning, then, may provide a parsimonious account of these differences between speech and limb control. Unlike the recombinable and highly overlearned syllabic plans in speech, there are unlikely to be preestablished movement plans for simple arm reaches in a virtual environment. Without separate representations of each independent movement subcomponent in reaching, there would be no conflict between the first movement (canceled-out adaptation) and the entire movement sequence (differential adaptation) that would lead to reduced adaptation. In this sense, our results not only extend findings in limb control to speech, they also highlight the widely established hierarchical nature of speech production.

Although other explanations for reduced adaptation in multisyllabic words are possible, we believe they are unlikely. One conceivable alternative explanation is that the difference between monosyllabic and multisyllabic words may be related to increased exposure to the perturbation due to longer vowels in monosyllables. However, several points argue against this interpretation. First, recent work in reaching showed little difference in adaptation when directly comparing short and long feedback durations presented during movement, as is the case for speech (Wang et al., 2024). Second, although vowel duration and syllable count were highly collinear across experiments, the binary variable of mono- versus multisyllabic targets was more predictive of the magnitude of differential adaptation across experiments than duration. Finally, a duration-based account would incorrectly predict a difference in adaptation size between Experiments 2 and 3, given the substantially shorter durations in the latter, yet no such difference was observed. In all, vowel duration seems insufficient to explain the differential adaptation size differences across the three experiments, particularly the mono- versus multisyllabic distinction.

A second alternative explanation is the effects of the lexical neighborhood: The altered auditory feedback resulted in target words sounding similar to real words in Experiment 1 (e.g., *bid, bad*), but not in Experiments 2 and 3 (e.g., *padigree*, *pidigree*). Previous work has suggested that perturbations of a real word that result in a nonword may have reduced adaptation size compared with perturbations that result in a real word, while the rate of adaptation is unaffected (Bourguignon et al., 2014). While modulatory effects of lexicality could thus potentially explain the reduced magnitude of differential adaptation in our multisyllabic words, they cannot explain the slower learning rate. In other words, lexicality may play a role but cannot be the sole driver of all mono- versus multisyllabic differences in our study.

A final possibility is that the reduced adaptation in Experiments 2 and 3 results not from competition between word- and syllable-level planning but rather from the slow emergence of novel, word-specific motor plans triggered by the repeated production of a limited set of stimulus words. Although word-level motor planning is included in some models of speech motor control, it is generally restricted to high-frequency words (e.g., Tourville & Guenther, 2011), forming prespecified selection units akin to frequent syllables in psycholinguistic models of speech production (e.g., Levelt, 2001). On the one hand, our results could suggest that all words form independent planning units regardless of their lexical frequency, given the presence of relatively low-frequency words of our multisyllabic stimuli (e.g., *pedicure*). On the other hand, if the repeated production of a limited set of words in Experiments 2 and 3 resulted in the ad hoc creation of new, word-level plans, learning would take longer to emerge but would eventually reach the same level as for monosyllabic words, as the word-level plans would replace, rather than compete with, syllable-level plans. Crucially, the adaptation in the multisyllabic words in Experiments 2 and 3 did not reach a plateau by the end of the hold phase as in the monosyllabic words; it is, therefore, unclear whether the adaptation size would always be less in multisyllabic words than in monosyllables or if the magnitude difference between the two would disappear with further perturbation exposure. Future research with extended exposure in the hold phase is needed to resolve this question. Regardless, our results still suggest that multisyllabic contexts, provided here by separate words, can serve as separate motor plans for otherwise identical syllables.

Theoretical Implications

By revealing the influence of multisyllabic (thus likely word-level) representations on the motor plan, our results have broad implications for our understanding of speech production. First and foremost, there is a long-standing fundamental debate regarding the information flow dynamics in speech production, especially regarding how word-level representation influences form preparation. Proposals range from fully discrete and feedforward connections (Levelt, 1992; segregated word and syllable activations), through limited interactivity (Goldrick et al., 2006; permitting word and syllable and coactivation), to highly interactive (Strijkers, 2016; parallel activations of syllable and word). The coactivation of words and syllables during motor planning, necessary to explain our results, is consistent with a large body of evidence supporting cascading rather than discrete activation between lexical and sublexical units (e.g., Alderete et al., 2021; Goldrick & Blumstein, 2006; Goldrick et al., 2011; Kurowski & Blumstein, 2016; Nozari & Dell, 2009; Oppenheim & Dell, 2008, 2010). However, existing models of word production typically involve abstract phonological sublexical units and, as such, do not make explicit claims about speech motor control. While the continued influence of word-level representations throughout speech articulation has some existing empirical support (see Dell et al., 2014a, 2014b; Goldrick, 2011, for more detailed discussions), it is also possible that word-level representation only cascades until the end of psycholinguistic preparation while still yielding distinct downstream acoustic differences. By using a novel assay of differential adaptation to specifically target speech motor planning, our study provides the critical missing link between word-level representations and articulation, suggesting that interactivity is not restricted to psycholinguistic preparation of abstract linguistic units such as (syllabified) phonemes but operates throughout the entire speech production process, including during speech motor planning.

Our finding that word context influences syllabic motor planning connects current models of spoken word production, on the one hand, and speech motor control, on the other. The link between psycholinguistic preparation (involving largely abstract linguistic units) and articulatory production is frequently modeled as a handoff or feedforward activation of syllabic motor plans (e.g., Dell, 1986; Hickok, 2012; Levelt, 2001; Tourville & Guenther, 2011).

Our results reveal that the interface of psycholinguistic preparation and speech motor control is more complicated and more than a simple feedforward activation of syllables: A multisyllabic/word representation remains present through articulatory planning.

Further, our results provide insights into the mental representation of words. The differential adaptation in Experiments 2 and 3 resulted in the same syllable being produced differently in different multisyllabic words, suggesting that the form representation of words involves fine-grained phonetic details, beyond simple concatenations of syllables. However, the reduced adaptation in the multisyllabic words compared with monosyllables, likely reflecting a conflict between whole-word and syllable-level planning, implies a shared (potentially abstract) syllable-level plan. Together, these findings support recent hybrid models of word representation that incorporate abstract and episodic representations (see Goldrick & Cole, 2023, for a review) rather than purely episodic accounts such as early versions of the exemplar theory (e.g., Pierrehumbert, 2001) or purely abstract cognitivist accounts (e.g., Dell, 1986).

Constraints on Generality

Finally, there are constraints on generality. Although our results provide evidence supporting suprasyllabic motor planning, there are several unanswered questions regarding the generalizability of this conclusion. As discussed above, our experiments provided direct evidence that speech production involves suprasyllabic motor plans, but the precise nature of these plans is still unclear. Although we believe our results are likely explained by word-level plans given the extensive literature on the relevance of words for both psycholinguistic preparation and speech motor planning, future research is needed to verify if these suprasyllabic motor plans are restricted to words or, potentially, would also be found for syllable sequences across word boundaries (e.g., metrical feet, phrases, sentences). Additionally, the study tested native English speakers based in the upper midwest of the United States; it remains to be tested whether the key findings still hold in other speaker groups, especially speakers of languages with drastically different vowel inventory and syllable structure (e.g., Japanese).

Conclusion

We found differential adaptation to opposing perturbations in monosyllabic and multisyllabic words with shared segmental content. Critically, the size of this differential adaptation was reduced in the multisyllabic words relative to the monosyllables, implying an interaction of (a) syllable-level and (b) suprasyllabic or word-level motor planning. These results have broad implications for our understanding of speech production. For the cognitive representations of speech, our results align with recent hybrid solutions that include abstract and episodic representations (e.g., Pierrehumbert, 2016). For psycholinguistic preparation, our results support models permitting word–syllable coactivation (e.g., Goldrick et al., 2006; Strijkers, 2016). For speech motor control, our results indicate that motor plans are not restricted to syllables and high-frequency words but may also include word-specific planning for even low-frequency words. Broadly, our results suggest two possibilities: on the one hand, the hand-off between psycholinguistic preparation and speech motor control may involve both words and syllables; alternatively, the integration of these two stages of speech production may be tighter than previously thought, implying that even higher level aspects of linguistic planning (e.g., semantic information) may also modulate speech motor planning. Future work is needed to resolve these questions. Overall, our study showcases how sensorimotor adaptation holds promise for investigating the dynamics of the speech production process, especially the interface of psycholinguistic preparation and speech motor control.

References

Alderete, J., Baese-Berk, M., Leung, K., & Goldrick, M. (2021). Cascading activation in phonological planning and articulation: Evidence from spontaneous speech errors. *Cognition*, *210*, Article 104577. https://doi.org/10.1016/j.cognition.2020.104577

Appelbaum, M., Cooper, H., Kline, R. B., Mayo-Wilson, E., Nezu, A. M., & Rao, S. M. (2018). Journal article reporting standards for quantitative research in psychology: The APA Publications and Communications Board task force report. *American Psychologist*, *73*(1), 3–25. https://doi.org/10.1037/amp0000191

Baayen, H., Vasishth, S., Kliegl, R., & Bates, D. (2017). The cave of shadows: Addressing the human factor with generalized additive mixed models. *Journal of Memory and Language*, *94*, 206–234. https://doi.org/10.1016/j.jml.2016.11.006

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., Dai, B., Scheipl, F., Grothendieck, G., Green, P., Fox, J., Bauer, A., & Krivitsky, P. N. (2023). *lme4: Linear mixed-effects models using "Eigen" and S4* (R package, Version 1.1-34) [Computer software]. https://CRAN.R-project.org/package=lme4

Boersma, P., & Weenink, D. (2023). *Praat: Doing phonetics by computer* (Version 6.3.18) [Computer software]. https://www.praat.org/

Bourguignon, N. J., Baum, S. R., & Shiller, D. M. (2014). Lexical-perceptual integration influences sensorimotor adaptation in speech. *Frontiers in Human Neuroscience*, *8*, Article 208. https://doi.org/10.3389/fnhum.2014.00208

Cai, S., Boucek, M., Ghosh, S. S., Guenther, F. H., & Perkell, J. S. (2008). A system for online dynamic perturbation of formant trajectories and results from perturbations of the Mandarin triphthong /iau/. *Proceedings of the 8th ISSP* (pp. 65–68).

Caudrelier, T., Schwartz, J.-L., Perrier, P., Gerber, S., & Rochet-Capellan, A. (2018). Transfer of learning: What does it tell us about speech production units? *Journal of Speech, Language, and Hearing Research*, *61*(7), 1613–1625. https://doi.org/10.1044/2018_JSLHR-S-17-0130

Cholin, J., Dell, G. S., & Levelt, W. J. M. (2011). Planning and articulation in incremental word production: Syllable-frequency effects in English. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(1), 109–122. https://doi.org/10.1037/a0021322

Cholin, J., Levelt, W. J. M., & Schiller, N. O. (2006). Effects of syllable frequency in speech production. *Cognition*, *99*(2), 205–235. https://doi.org/10.1016/j.cognition.2005.01.009

Correll, J., Mellinger, C., & Pedersen, E. J. (2022). Flexible approaches for estimating partial eta squared in mixed-effects models with crossed random factors. *Behavior Research Methods*, *54*(4), 1626–1642. https://doi.org/10.3758/s13428-021-01687-2

de Rosario-Martinez, H., Fox, J., R Core Team, & Phil, C. (2015). *phia: Post-hoc interaction analysis* (Version 0.2-1) [Computer software]. https://cran.r-project.org/web/packages/phia/index.html

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, *93*(3), 283–321. https://doi.org/10.1037/0033-295X.93.3.283

Dell, G. S., Nozari, N., & Oppenheim, G. M. (2014a). Word production: Behavioral and computational considerations. In M. Goldrick, V. Ferreira, & M. Miozzo (Eds.), *The Oxford handbook of language production* (pp. 88–104). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199735471.013.014

Dell, G. S., Nozari, N., & Oppenheim, G. M. (2014b). Lexical access: Behavioral and computational considerations. In M. Goldrick, M. Miozzo, & V. S. Ferreira (Eds.), *The Oxford handbook of language production* (pp. 88–104). Oxford University Press.

Goldrick, M. (2011). Linking speech errors and generative phonological theory: Speech errors and phonological theory. *Language and Linguistics Compass*, *5*(6), 397–412. https://doi.org/10.1111/j.1749-818X.2011.00282.x

Goldrick, M., & Blumstein, S. E. (2006). Cascading activation from phonological planning to articulatory processes: Evidence from tongue twisters. *Language and Cognitive Processes*, *21*(6), 649–683. https://doi.org/10.1080/01690960500181332

Goldrick, M., & Cole, J. (2023). Advancement of phonetics in the 21st century: Exemplar models of speech production. *Journal of Phonetics*, *99*, Article 101254. https://doi.org/10.1016/j.wocn.2023.101254

Goldrick, M., Miozzo, M., & Ferreira, V. S. (2006). Limited interaction in speech production: Chronometric, speech error, and neuropsychological evidence. *Language and Cognitive Processes*, *21*(7–8), 817–855. https://doi.org/10.1080/01690960600824112

Goldrick, M., Ross Baker, H., Murphy, A., & Baese-Berk, M. (2011). Interaction and representational integration: Evidence from speech errors. *Cognition*, *121*(1), 58–72. https://doi.org/10.1016/j.cognition.2011.05.006

Guenther, F. H. (2016). *Neural control of speech*. MIT Press. https://doi.org/10.7551/mitpress/10471.001.0001

Heald, J. B., Lengyel, M., & Wolpert, D. M. (2023). Contextual inference in learning and memory. *Trends in Cognitive Sciences*, *27*(1), 43–64. https://doi.org/10.1016/j.tics.2022.10.004

Hickok, G. (2012). The cortical organization of speech processing: Feedback control and predictive coding the context of a dual-stream model. *Journal of Communication Disorders*, *45*(6), 393–402. https://doi.org/10.1016/j.jcomdis.2012.06.004

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics, Theory and Applications*, *6*(2), 65–70.

Howard, I. S., Wolpert, D. M., & Franklin, D. W. (2013). The effect of contextual cues on the encoding of motor memories. *Journal of Neurophysiology*, *109*(10), 2632–2644. https://doi.org/10.1152/jn.00773.2012

Howard, I. S., Wolpert, D. M., & Franklin, D. W. (2015). The value of the follow-through derives from motor learning depending on future actions. *Current Biology*, *25*(3), 397–401. https://doi.org/10.1016/j.cub.2014.12.037

Jones, J. A., & Munhall, K. G. (2000). Perceptual calibration of F0 production: Evidence from feedback perturbation. *Journal of the Acoustical Society of America*, *108*(3), 1246–1251. https://doi.org/10.1121/1.1288414

Kim, K. S., Wang, H., & Max, L. (2020). It's about time: Minimizing hardware and software latencies in speech research with real-time auditory feedback. *Journal of Speech, Language, and Hearing Research*, *63*(8), 2522–2534. https://doi.org/10.1044/2020_JSLHR-19-00419

Krause, J. C., & Braida, L. D. (2004). Acoustic properties of naturally produced clear speech at normal speaking rates. *The Journal of the Acoustical Society of America*, *115*(1), 362–378. https://doi.org/10.1121/1.1635842

Kurowski, K., & Blumstein, S. E. (2016). Phonetic basis of phonemic paraphasias in aphasia: Evidence for cascading activation. *Cortex*, *75*, 193–203. https://doi.org/10.1016/j.cortex.2015.12.005

Kuznetsova, A., Brockhoff, P. B., Christensen, R. H. B., & Jensen, S. P. (2020). *lmerTest: Tests in linear mixed effects models* (R, Version 3.1-3) [Computer software]. https://CRAN.R-project.org/package=lmerTest

Levelt, W. J. M. (1992). Accessing words in speech production: Stages, processes and representations. *Cognition*, *42*(1–3), 1–22. https://doi.org/10.1016/0010-0277(92)90038-J

Levelt, W. J. M. (1999). Models of word production. *Trends in Cognitive Sciences*, *3*(6), 223–232. https://doi.org/10.1016/S1364-6613(99)01319-4

Levelt, W. J. M. (2001). Spoken word production: A theory of lexical access. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(23), 13464–13471. https://doi.org/10.1073/pnas.231459498

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In W. J. Hardcastle & A. Marchal (Eds.), *Speech production and speech modelling* (pp. 403–439). Springer Netherlands. https://doi.org/10.1007/978-94-009-2037-8_16

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing type I error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305–315. https://doi.org/10.1016/j.jml.2017.01.001

Munhall, K. G., MacDonald, E. N., Byrne, S. K., & Johnsrude, I. (2009). Talkers alter vowel production in response to real-time formant perturbation even when instructed not to compensate. *The Journal of the Acoustical Society of America*, *125*(1), 384–390. https://doi.org/10.1121/1.3035829

Nakagawa, S., Johnson, P. C. D., & Schielzeth, H. (2017). The coefficient of determination R^2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *Journal of the Royal Society, Interface*, *14*(134), Article 20170213. https://doi.org/10.1098/rsif.2017.0213

Niziolek, C. A. (2014). When bed goes bad: How the brain can fix mistakes in speech while they happen. *Frontiers for Young Minds*, *2*, Article 2. https://doi.org/10.3389/frym.2014.00002

Niziolek, C. A., & Houde, J. F. (2015). *wave_viewer: First release* [Computer software]. Zenodo. https://doi.org/10.5281/zenodo.13839

Niziolek, C. A., & Parrell, B. (2021). Responses to auditory feedback manipulations in speech may be affected by previous exposure to auditory errors. *Journal of Speech, Language, and Hearing Research*, *64*(6S), 2169–2181. https://doi.org/10.1044/2020_JSLHR-20-00263

Noiray, A., Cathiard, M.-A., Ménard, L., & Abry, C. (2011). Test of the movement expansion model: Anticipatory vowel lip protrusion and constriction in French and English speakers. *The Journal of the Acoustical Society of America*, *129*(1), 340–349. https://doi.org/10.1121/1.3518452

Nozari, N., & Dell, G. S. (2009). More on lexical bias: How efficient can a "lexical editor" be? *Journal of Memory and Language*, *60*(2), 291–307. https://doi.org/10.1016/j.jml.2008.09.006

Oppenheim, G. M., & Dell, G. S. (2008). Inner speech slips exhibit lexical bias, but not the phonemic similarity effect. *Cognition*, *106*(1), 528–537. https://doi.org/10.1016/j.cognition.2007.02.006

Oppenheim, G. M., & Dell, G. S. (2010). Motor movement matters: The flexible abstractness of inner speech. *Memory & Cognition*, *38*(8), 1147–1160. https://doi.org/10.3758/MC.38.8.1147

Orwant, J., & Brockman, W. (2019). *Google Ngram viewer* [Computer software]. https://books.google.com/ngrams/

Osu, R., Hirai, S., Yoshioka, T., & Kawato, M. (2004). Random presentation enables subjects to adapt to two opposing forces on the hand. *Nature Neuroscience*, *7*(2), 111–112. https://doi.org/10.1038/nn1184

Parrell, B., Lammert, A. C., Ciccarelli, G., & Quatieri, T. F. (2019). Current models of speech motor control: A control-theoretic overview of architectures and properties. *The Journal of the Acoustical Society of America*, *145*(3), 1456–1481. https://doi.org/10.1121/1.5092807

Parrell, B., & Niziolek, C. A. (2021). Increased speech contrast induced by sensorimotor adaptation to a nonuniform auditory perturbation. *Journal of Neurophysiology*, *125*(2), 638–647. https://doi.org/10.1152/jn.00466.2020

Perkell, J. S., & Matthies, M. L. (1992). Temporal measures of anticipatory labial coarticulation for the vowel/u/: Within- and cross-subject variability. *Journal of the Acoustical Society of America*, *91*(5), 2911–2925. https://doi.org/10.1121/1.403778

Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency, lenition and contrast. In J. L. Bybee & P. J. Hopper (Eds.), *Frequency and the emergence of linguistic structure* (pp. 137–157). John Benjamins Publishing. https://doi.org/10.1075/tsl.45.08pie

Pierrehumbert, J. B. (2016). Phonological representation: Beyond abstract versus episodic. *Annual Review of Linguistics*, *2*(1), 33–52. https://doi.org/10.1146/annurev-linguistics-030514-125050

Pluymaekers, M., Ernestus, M., & Baayen, R. H. (2005). Lexical frequency and acoustic reduction in spoken Dutch. *The Journal of the Acoustical Society of America*, *118*(4), 2561–2569. https://doi.org/10.1121/1.2011150

Pouplier, M. (2007). Tongue kinematics during utterances elicited with the SLIP technique. *Language and Speech*, *50*(3), 311–341. https://doi.org/10.1177/00238309070500030201

R Core Team. (2023). *R: A language and environment for statistical computing* (Version 4.3.1) [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Recasens, D. (2018). *Coarticulation*. Oxford Research Encyclopedia of Linguistics. https://doi.org/10.1093/acrefore/9780199384655.013.416

Rochet-Capellan, A., & Ostry, D. J. (2011). Simultaneous acquisition of multiple auditory–motor transformations in speech. *The Journal of Neuroscience*, *31*(7), 2657–2662. https://doi.org/10.1523/JNEUROSCI.6020-10.2011

Schween, R., Taube, W., Gollhofer, A., & Leukel, C. (2014). Online and post-trial feedback differentially affect implicit adaptation to a visuomotor rotation. *Experimental Brain Research*, *232*(9), 3007–3013. https://doi.org/10.1007/s00221-014-3992-z

Shattuck-Hufnagel, S., & Klatt, D. H. (1979). The limited use of distinctive features and markedness in speech production: Evidence from speech error data. *Journal of Verbal Learning and Verbal Behavior*, *18*(1), 41–55. https://doi.org/10.1016/S0022-5371(79)90554-1

Sheahan, H. R., Franklin, D. W., & Wolpert, D. M. (2016). Motor planning, not execution, separates motor memories. *Neuron*, *92*(4), 773–779. https://doi.org/10.1016/j.neuron.2016.10.017

Stevens, S. S., Volkmann, J., & Newman, E. B. (1937). A scale for the measurement of the psychological magnitude pitch. *Journal of the Acoustical Society of America*, *8*(3), 185–190. https://doi.org/10.1121/1.1915893

Strijkers, K. (2016). A neural assembly-based view on word production: The bilingual test case. *Language Learning*, *66*(S2), 92–131. https://doi.org/10.1111/lang.12191

Tourville, J. A., Cai, S., & Guenther, F. (2013). Exploring auditory–motor interactions in normal and disordered speech. *Proceedings of Meetings on Acoustics Acoustical Society of America*, *19*, Article 060180. https://doi.org/10.1121/1.4800684

Tourville, J. A., & Guenther, F. H. (2011). The DIVA model: A neural theory of speech acquisition and production. *Language and Cognitive Processes*, *26*(7), 952–981. https://doi.org/10.1080/01690960903498424

Umeda, N. (1975). Vowel duration in American English. *The Journal of the Acoustical Society of America*, *58*(2), 434–445. https://doi.org/10.1121/1.380688

Van der Merwe, A. (2021). New perspectives on speech motor planning and programming in the context of the four-level model and its implications for understanding the pathophysiology underlying apraxia of speech and other motor speech disorders. *Aphasiology*, *35*(4), 397–423. https://doi.org/10.1080/02687038.2020.1765306

van Rij, J., Wieling, M., Baayen, R. H., & van Rijn, H. (2022). *itsadug: Interpreting time series and autocorrelated data using GAMMs* (Version 2.4.1) [Computer software]. https://cran.r-project.org/web/packages/itsadug/index.html

Voeten, C. C. (2023). *buildmer: Stepwise elimination and term reordering for mixed-effects regression* (Version 2.9) [Computer software]. https://CRAN.R-project.org/package=buildmer

Wang, T., Avraham, G., Tsay, J. S., Thummala, T., & Ivry, R. B. (2024). Advanced feedback enhances sensorimotor adaptation. *Current Biology*, *34*(5), 1076–1085.e5. https://doi.org/10.1016/j.cub.2024.01.073

Whalen, D. H. (1990). Coarticulation is largely planned. *Journal of Phonetics*, *18*(1), 3–35. https://doi.org/10.1016/S0095-4470(19)30356-0

Wieling, M. (2018). Analyzing dynamic phonetic data using generalized additive mixed modeling: A tutorial focusing on articulatory differences between L1 and L2 speakers of English. *Journal of Phonetics*, *70*, 86–116. https://doi.org/10.1016/j.wocn.2018.03.002

Wood, S. (2023). *mgcv: Mixed GAM computation vehicle with automatic smoothness estimation* (Version 1.9-0) [Computer software]. https://cran.r-project.org/web/packages/mgcv/index.html

Received January 15, 2024
Revision received January 8, 2025
Accepted January 19, 2025 ■