

Transfer of Sensorimotor Adaptation Reveals Independent Prosodic Representation After Segment–Prosody Coordination in Speech Production

Yuyu Zeng¹, Benjamin Parrell^{1, 2}, and Caroline A. Niziolek^{1, 2}

¹ Waisman Center, University of Wisconsin–Madison

² Department of Communication Sciences and Disorders, University of Wisconsin–Madison

In speech production, there is a long-standing debate regarding whether word-level prosodic structure has an independent representation separable from segments. Using a novel assay of sensorimotor adaptation, we observe straightforward evidence supporting an independent prosodic structure even after its coordination with the segments. Participants exposed to opposing formant perturbations applied to the two syllables of a single trained word (e.g., “bedhead” → “bidhad”) adapted separately to counteract the syllable-specific perturbations (i.e., producing “badhid”) and, critically, transferred the learned adaptation to untrained words based purely on shared prosodic structure (words with the same prosody but novel syllables, e.g., producing “breastfed” like “brastfid”). This evidence for an independent word-level prosodic representation, which has been hard to detect in previous studies, highlights the usefulness of sensorimotor adaptation as a tool for language production research.

Keywords: speech motor control, prosody, segments, sensorimotor adaptation, altered auditory feedback

Supplemental materials: <https://doi.org/10.1037/xlm0001551.supp>

Running speech consists of both the articulation of individual speech segments (vowels and consonants) and the production of speech prosody (e.g., pitch contour, stress, and rhythm). For example, differences in sentence-level prosody can create highly different meanings (e.g., “What’s that ahead in the road?” vs. “What’s that, a HEAD in the ROAD?” examples from [Speer & Blodgett, 2006](#)), and word-level prosody distinguishes the noun *PERmit* from the verb *perMIT* (capital letters indicate primary stress). There is continued debate regarding the relationship between segmental and prosodic planning during speech production: Some models propose that prosody is planned first, integrating segmental and prosodic planning from the outset, with segments filling in the slots in a frame created by the initial prosodic planning (the prosody-first account; e.g., [Keating & Shattuck-Hufnagel, 2002](#); [Turk & Shattuck-Hufnagel, 2020](#)), but others advocate for the early planning of segments, followed by prosody being superimposed on already planned segments (the segment-first account; e.g., [Levett, 1989](#); [Levett et al., 1999](#)).

Prosodic changes and segmental articulation co-occur in time in running speech, so there must be a mechanism that coordinates their planning in the temporal domain, regardless of the initial planning

order. Crucially, the prosody-first and segment-first accounts differ not only in the order in which planning occurs but also in their view of the independence of prosodic representation after this coordination: While the prosody-first account was strongly influenced by speech error data suggesting the independence of prosody (e.g., [Shattuck-Hufnagel, 1992](#)), the segment-first account was more influenced by priming experiments that favor a unified representation of segments and prosody (e.g., [Myers & Watson, 2019](#); [Roelofs & Meyer, 1998](#)), with no separate representation of prosody (independent from segments). For example, [Myers and Watson \(2019\)](#) asked speakers to describe objects using formulaic sentences (*The X shrinks. The Y flashes.*); the critical nouns (*X* and *Y*) had different prime–target relationships in the first syllables, overlapping in segmental contents (*canTEEN* vs. *CANdle*), lexical stress (*TRUMPet* vs. *CANdle*), or both features (*CANdy* vs. *CANdle*). They found that target word duration was the longest (suggesting the largest effect of inhibitory priming on word planning) when a prime–target pair overlapped in both segments and lexical stress placement; inhibitory priming was reduced when only the segments were shared and was absent when only the lexical stress was shared. These results suggest that after the segment–prosody coordination,

This article was published Online First December 8, 2025.

Heather Ferguson served as action editor.

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

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

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

Benjamin Parrell and Caroline A. Niziolek are equal contributors. All data are publicly available at the Open Science Framework (<https://osf.io/d4t6s/>).

Caroline A. Niziolek and Benjamin Parrell received funding from the National Science Foundation, Division of Behavioral and Cognitive Sciences (Grant 2120506).

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. Benjamin Parrell 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. 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.

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

the segments maintain a quasi-independent representation where priming is possible across prosodic contexts, but that there is no independent prosodic representation separable from segments (at least for lexical stress).

However, there is evidence that priming may not provide an accurate diagnostic of prosodic structure. For example, in Mandarin Chinese, priming based on lexical tone is both weaker and less consistently found than priming based on segments (Zeng, 2022), even though lexical tone has a functional load higher or equivalent to that of vowels (Oh et al., 2015; Surendran & Levow, 2004); that is, tone contrasts play a substantially greater role in distinguishing meaning in Mandarin than lexical stress contrasts do in English (functional load: 4.13 for Mandarin lexical tone vs. 0.005 for English lexical stress, according to Oh et al., 2015). If detecting prosodic effects in Chinese (lexical tone) is difficult, one can reasonably assume that detecting prosodic effects in English (lexical stress) is even more challenging. In other words, priming may not be effective at eliciting prosodic effects, and the weakness/absence of prosodic priming may reflect the inherent limitation of the priming procedure rather than a lack of independent prosodic representation. Here, we propose using sensorimotor adaptation as a more direct assay of the independence of prosodic representation after the coordination of prosodic and segmental planning. Prosodic effects have been hard to detect, but the sensorimotor adaptation procedure has relatively high task sensitivity. For example, a group size of 10 is sufficient to observe robust vowel production changes in English sentences that vary markedly in length, vocabulary, and grammar (Lametti et al., 2018). More importantly, previous research indicates this procedure is sensitive to prosodic manipulations, including lexical stress (Bakst & Niziolek, 2021) and lexical tone (see Tang, 2024, for a recent review).

In a typical speech sensorimotor adaptation study, the acoustic signals of speech are modified by shifting vowel formants (the resonant frequencies of the vocal tract) in near real-time. For example, speakers producing *bed* may hear something more like *bad* (Figure 1A) and consequently adapt their productions to oppose the auditory perturbation (e.g., by shifting their produced vowel formants toward those of *bid*). Examining how this adaptation transfers to untrained items has provided insights into the role of prosodic structure in speech. For example, French speakers who adapted to a vowel formant perturbation in monosyllabic *bé* showed incomplete transfer to the same syllable in the multisyllabic words *bépé* and *bébé*, with larger (yet still incomplete) transfer to word-initial than word-final *bé* syllables (Caudrelier et al., 2018). While this result suggests that prosodic structure modulates the transfer of learned adaptation in a particular segmental sequence, it remains unclear whether the prosodic structure *alone* can enable adaptation transfer in a similar manner as segmental content, as predicted if prosodic structure retains an independent representation during production. Here, we test this prosodic transfer to directly evaluate the independent representation of prosody. Given the broad scope of speech prosody (Gussenhoven & Chen, 2021), we focus here on the same, relatively well-studied aspect of prosody as in Myers and Watson (2019): the word-level prosodic frame (Shattuck-Hufnagel, 2015),¹ which specifies syllable positions and lexical stress placement in English.

To test the independence of the prosodic frame after its coordination with segments, we applied opposing formant perturbations

to the two syllables of the word *bedhead* (Figure 1C, left): Because these syllables share the same vowel /e/, any adaptation based purely on vowel identity should sum to zero. Although speakers can adapt to opposing perturbations in separate monosyllabic words sharing a single vowel (e.g., *bed* and *head*; Rochet-Capellan & Ostry, 2011), it is currently unknown whether this ability extends to different syllables within the same word. After first confirming that speakers can indeed adapt separately in the two syllables of *bedhead* (Experiment 1), we then tested how adaptation transfers to words with the same stress placement but differing in their segmental content (e.g., *breastfed*; Figure 1C, right; Experiment 2). Because these transfer words were not themselves perturbed, any changes in their production must come from some shared structural representation with the trained word *bedhead*. There are several possibilities for how learning may transfer based on segmental or prosodic structure; below, we lay out these alternatives and what they would suggest about the representation of prosodic structure.

There are two simple predictions, though both are unlikely given previous work in speech sensorimotor adaptation and prosodic/segmental planning (e.g., Caudrelier et al., 2018; see Shattuck-Hufnagel, 2020, for a recent overview). On the one hand, if produced words contain an independent representation of segments entirely free from the prosodic structure, we should see complete adaptation transfer in the trained syllables, regardless of the prosodic structure of the novel word (e.g., equivalent changes for the syllable “bed” in *bedhead*, *bedspread*, *deathbed*, and *bed*). On the other hand, if produced words contain an independent representation of prosody, free from the influence of segments, we should see complete adaptation transfer in all disyllabic words with the same prosodic frame (e.g., equivalent changes for both syllables in *bedhead* and *breastfed*).

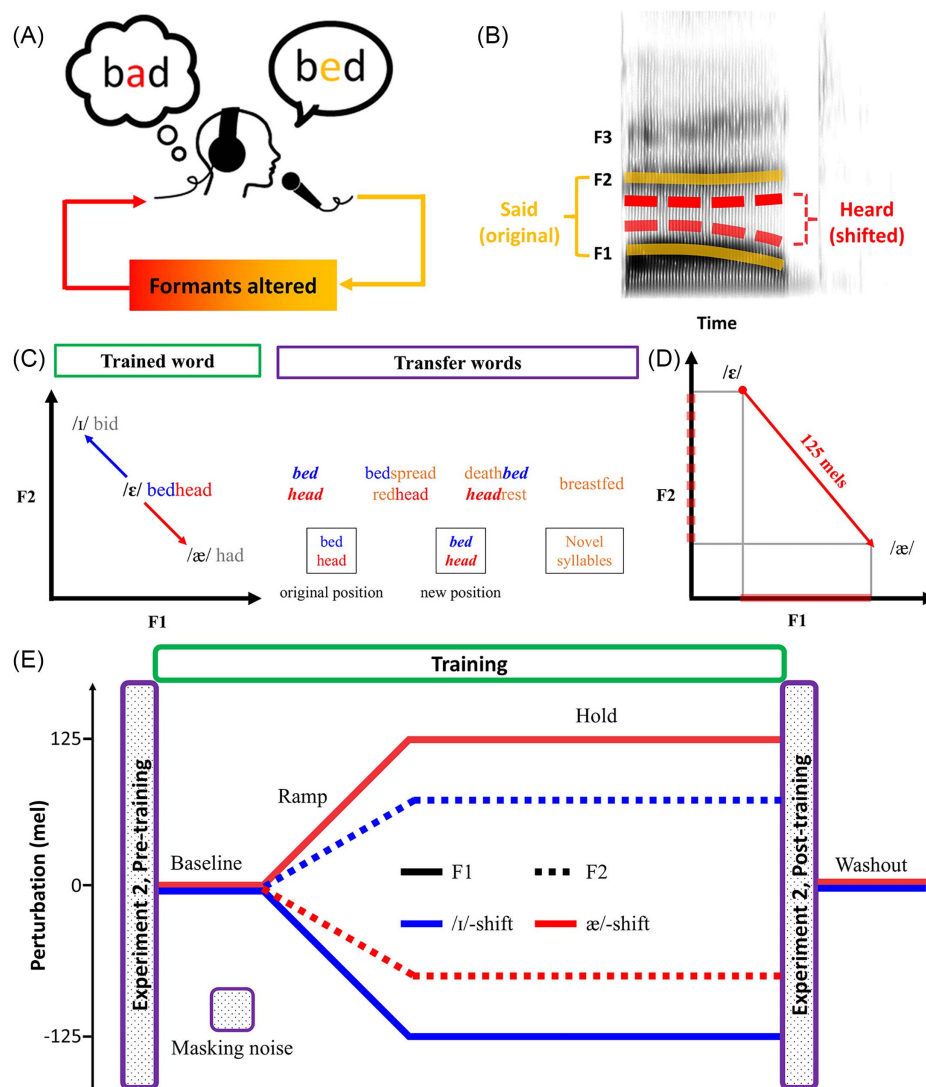
More likely are predictions derived from the segment-first and prosody-first models that embrace segment–prosody coordination: Both models predict that segmental learning should transfer across prosodic positions, but this transfer should be reduced in the new prosodic positions compared to the original prosodic positions (e.g., larger changes for the syllable “bed” in *bedhead* than *deathbed*). However, only the prosody-first model, which assumes an independent prosodic representation, predicts that adaptation should (partially) transfer to the vowel in novel words based solely on their prosodic position (e.g., from *bedhead* to *breastfed*), resulting from the persistence of an independent prosodic representation after its coordination with segments. In contrast, the segment-first account, which suggests that prosodic structure does not have an independent representation, predicts no adaptation transfer based on shared prosodic structure (e.g., no transfer from *bedhead* to *breastfed*).

Method

All experiments and data analyses can be reconstructed using the shared code at the Open Science Framework (Zeng et al., 2025b; <https://osf.io/d4t6s/>); this link also contains the Supplemental Materials. The

¹ When individual words are considered, the term “metrical frame” is often used synonymously with prosodic frame (e.g., Levelt, 1992; Roelofs & Meyer, 1998). In this article, we only use “prosodic frame” to avoid confusion.

Figure 1
Sensorimotor Adaptation Procedure and Experiment Design



Note. (A) A speaker's produced speech is altered and played back via headphones in near real-time (from "bed" to "bad"). Adapted from "When Bed Goes Bad: How the Brain Can Fix Mistakes in Speech While They Happen," by C. Niziolek, 2014, *Frontiers for Young Minds*, 2(2) <https://doi.org/10.3389/frym.2014.00002>. Copyright 2014 by Niziolek. Adapted with permission. (B) Spectrogram of the speech signal for "bed" and the formant alterations. The darker bands are energy concentrations at specific frequencies, corresponding to vowel formants (F_1 , F_2 , and F_3), the primary acoustic correlates of vowels. In the example, F_1 and F_2 are shifted to sound more like "bad." (C) The trained word (left) and transfer words (right) in Experiment 2, with one example of perturbations in an F_1 - F_2 space (shifting the initial syllable toward /t/, but shifting the second syllable toward /æ/; counterbalanced across participants). (D) A schematic of how a 125-mel perturbation toward /æ/ is projected onto F_1 and F_2 dimensions. (E) A schematic of the overall procedure, including the relative magnitude of F_1 and F_2 perturbations. See the online article for the color version of this figure.

institutional review board at the University of Wisconsin–Madison approved all procedures.

Participants

Twenty speakers participated in Experiment 1 (15 female, four male, and one nonbinary; $M_{\text{age}} = 22.30$, $SD = 8.02$, range = 19–55).

Twenty speakers participated in Experiment 2 (18 female and two male; $M_{\text{age}} = 20.14$, $SD = 2.33$, range = 18–29). All participants were adult native English speakers, with some additionally speaking at least another language (nine in Experiment 1 and eight in Experiment 2). 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 250–4,000 Hz range). Participants were compensated either with course credit or a monetary payment.

Experiment Design

Experiment 1 establishes that English speakers adapt to opposing perturbations applied to the two syllables of the trained word *bedhead*, and Experiment 2 further examines how the learned adaptation transfers across segments and prosodic positions by having English speakers produce words before and after the training. The opposing perturbations were defined in an $F1$ – $F2$ space (Figure 1C, left), where one perturbation shifted toward /i/ (as in *bid* and *hid*), while the other perturbation shifted toward /æ/ (as in *bad* and *had*). The assignment of perturbation directions to the two syllables of *bedhead* was counterbalanced across participants in both experiments. Each speaker received individualized perturbations according to their vowel space organization (see the Supplemental Materials for details). The perturbation magnitude was fixed at 125 mels for both the /i/-shift and /æ/-shift, defined in an $F1$ – $F2$ space (Figure 1D).² Ideally, the /i/-shift would result in downward $F1$ perturbation and upward $F2$ perturbation, while the /æ/-shift would result in upward $F1$ perturbation and downward $F2$ perturbation (Figure 1C, left). When the perturbations were projected onto the $F1$ and $F2$ dimensions per participant, the two $F1$ perturbations were always in opposite directions, as expected. However, not all participants' $F2$ perturbations were opposing (see the Supplemental Materials for details): Only data from participants whose $F2$ perturbations were in opposite directions for /i/-shift and /æ/-shift were included in analyses of $F2$ adaptation (13 participants in Experiment 1 and 15 participants in Experiment 2). Across experiments and shift directions, the average absolute magnitudes of these perturbations applied to $F1$ and $F2$ were 107.62 mels and 56.40 mels, respectively (see the Supplemental Material for details).

Experiment 2 tested adaptation transfer to four types of words (Figure 1C, right): (a) the trained syllables as isolated monosyllabic words (*bed* and *head*); (b) the trained syllables in their original prosodic position, combined with another novel syllable (*bedspread* and *redhead*); (c) the trained syllables in the opposite prosodic position, combined with another novel syllable (*deathbed* and *headrest*); (d) novel syllables in both prosodic positions (*breastfed*). All syllables contained the same vowel /e/, and all disyllabic transfer words had the same prosodic frame as the trained word (primary stress on the initial syllable).

Experiment 1 only tested the trained word *bedhead*, consisting of four phases (Figure 1E): baseline (30 trials), ramp (30 trials), hold (150 trials), and washout (30 trials). No perturbation was present in the baseline and the washout phases. The perturbation magnitude increased linearly throughout the ramp phase, reaching a maximum of 125 mels, which was maintained during the hold phase. The angle of the formant perturbation was set individually for each participant, depending on their vowel space organization. A self-timed break was included every 30 trials.

Experiment 2 followed the same basic structure as Experiment 1, with the addition of a pretraining phase before the baseline phase and a posttraining phase after the hold phase (Figure 1E). In these two phases, untrained transfer words plus the trained word *bedhead* were presented in cycles, each containing one

repetition of each word. The order of words within each cycle was pseudorandomized, such that no two adjacent words were the same. Additionally, auditory feedback in these phases was masked by speech-shaped noise (~ 77 dB sound pressure level [SPL]) to prevent participants from using auditory feedback to adjust their speech online, enabling a cleaner measure of adaptation transfer from the trained word to the transfer words in the posttraining phase. Accordingly, there were six phases in Experiment 2: pretraining (10 cycles containing trained and transfer words), baseline (10 trials of *bedhead*), ramp (20 trials of *bedhead*), hold (200 trials of *bedhead*), posttraining (10 cycles containing trained and transfer words), and washout (10 trials of *bedhead*). The hold phase in Experiment 2 was extended relative to Experiment 1 to maximize learning and potential transfer, as we observed that adaptation had not yet reached a plateau in Experiment 1 after 150 trials.

Procedure and Equipment

Participants sat in front of a computer screen where one word appeared per trial. They were instructed to read these stimuli aloud as they appeared on the screen. On each trial, a target word appeared at the screen center for 1,400 ms (white text on a black background). The intertrial duration was 1,250 ms, with a 250 ms jitter. If the recorded speech signal in a trial did not meet a prespecified minimal intensity threshold, that trial was repeated (typically, this occurred when participants failed to produce any response). 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 $F1$ and $F2$) 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 playback was set to ~ 80 dB SPL and mixed with speech-shaped noise at ~ 60 dB SPL to mask air- and bone-conducted feedback; the intensity of speech playback varied with the intensity of produced speech. The speech was processed similarly through Audapter on all trials, regardless of whether a perturbation was applied. For trials with masking noise in Experiment 2 (pretraining and posttraining phases), only speech-shaped noise, without any speech signal, was played back at ~ 77 dB SPL. All stimulus presentation and data collection used MATLAB (The MathWorks, Inc.).

Before the main experiment, participants completed a brief calibration phase where *bid*, *bed*, and *bat* were each repeated 10 times (order randomized). This task achieved two goals: (a) to obtain a speaker-specific linear predictive coding (LPC) order for more accurate formant tracking and alteration in Audapter; (b) to calculate speaker-specific perturbation angles in an $F1$ – $F2$ space. 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.

² The magnitude of the formant perturbation was measured in mels, a perceptual scale across frequencies where each step along the mel scale is perceived as equally distant (Stevens et al., 1937).

Data Processing

The Montreal Forced Aligner (McAuliffe et al., 2017) automatically segmented recorded speech into phonemes; boundaries were checked and adjusted manually when the boundaries identified by the Montreal Forced Aligner were unsatisfactory. For each vowel, $F1$ and $F2$ measurements were obtained every 3 ms using Praat (Boersma & Weenink, 2023) via the wave_viewer package (C. A. Niziolek & Houde, 2015). Preemphasis values and linear predictive coding (LPC) order were set for each participant; errors in formant tracking (e.g., sudden jumps, tracking the wrong formants) were corrected with minimal adjustments to these parameters. Trials with unresolvable formant tracking errors or unintended productions (e.g., coughing, hesitations) were excluded (Experiment 1: $M = 0.53\%$, $SD = 1.49\%$, range = [0%, 6.67%]; Experiment 2: $M = 1.53\%$, $SD = 1.65\%$, range = [0%, 4.76%]).

For each trial, $F1$ and $F2$ measurements from 25% to 75% of a vowel were averaged and converted to the mel scale. To capture the $F1$ and $F2$ changes for each speaker, the average values were normalized per word per syllable: (a) To evaluate whether adaptation occurred, each syllable in *bedhead* was normalized relative to the mean $F1$ and $F2$ for that syllable in the last 10 baseline trials; (b) to evaluate the presence of adaptation transfer in each prosodic position, all syllables per word in the posttraining phase (including the transfer words and the trained word *bedhead*) were normalized relative to the mean $F1$ and $F2$ of that same syllable per word in the pretraining phase.

In order to verify that the disyllabic transfer words in Experiment 2 were produced with the same stress placement as the trained word *bedhead*, we also measured three types of acoustic cues commonly associated with English lexical stress: fundamental frequency (f_0), intensity, and duration (Lieberman, 1960; Plag et al., 2011). All measurements were obtained from the vocalic portion of each syllable in both the pretraining and posttraining phases. The f_0 was extracted using Praat (Boersma & Weenink, 2023), where the expected f_0 range was adjusted per speaker to avoid any pitch doubling or halving; the median f_0 of each syllable in each trial was used in the statistical analysis. Two intensity measurements were also extracted: the average and the maximum RMS (root-mean-square) intensity of the vocalic portion of each syllable.

Statistical Analysis

The primary statistical analysis was 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)³, where the reported p values were based on the lmerTest package (Kuznetsova et al., 2020; Version 3.1-3). In all analyses, the α level was 0.05. The best model was selected using the buildmer package (Voeten, 2023; Version 2.9), which systematically compares models differing in one term. To avoid inflating the Type I error rate (Barr et al., 2013), the initial formula (see below) included all possible interactions of the independent variables as fixed effects; the random effects included both participant and relevant grouping variables (such as word) as random intercepts, plus a full random-slope structure. 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 found; a backward-fitting procedure then

eliminated terms that did not contribute to model fit. In both forward-fitting and backward-fitting, the terms in each model were ordered according to their contribution to the overall fit (Matuschek et al., 2017) using the Bayesian information criterion (BIC), which balances fit and complexity. Specifically, the most important terms were the first to be included in forward-fitting, and the least important terms were the first to be eliminated in backward-fitting. The phia package (de Rosario-Martinez et al., 2015; Version 0.2-1) was used for post hoc comparisons, with the reported p values after the correction procedure in Holm (1979). As the proper way to estimate effect size for mixed-effects models is still under development (Correll et al., 2022), we report R^2 to represent the total variance explained by the selected model and the individual terms (Nakagawa et al., 2017). When relevant, 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.

For both experiments, we confirmed the presence of adaptation by examining the $F1$ and $F2$ changes ($\Delta F1$ and $\Delta F2$; normalized relative to the average of the last 10 baseline trials) in individual syllables in the trained word *bedhead* during training (baseline, ramp, hold, and washout phases). As our hypothesis relies critically on adaptation to opposing perturbations, and the / i /-shift and / $\ae$ /-shift in $F1$ - $F2$ space were not necessarily directly (or even roughly) opposing, depending on each participant's vowel space (see Supplemental Figures S1 and S2), $F1$ and $F2$ were analyzed separately as the projected perturbations on each dimension. The two shifts were in opposite directions for all participants in $F1$ and for a large subset of participants in $F2$. $F1$ and $F2$ data from the last 10 hold trials and the first 10 washout trials were analyzed. The initial formulae included the following independent variables: ShiftDirection (/ i /-shift vs. / $\ae$ /-shift), Phase (hold and washout), and ShiftOrder (whether the initial syllable received / i /-shift or / $\ae$ /-shift, accounting for potential order effects created by counterbalancing). Adaptation should surface as a significant difference between the / i /-shift and the / $\ae$ /-shift (a main effect of ShiftDirection). Whether specific perturbation conditions differed significantly from zero was also evaluated with additional t tests.

$$\Delta F1|\Delta F2 \sim \text{ShiftDirection} * \text{Phase} * \text{ShiftOrder} \\ + (\text{ShiftDirection} * \text{Phase} \\ * \text{ShiftOrder} | \text{Speaker}). \quad (1)$$

For Experiment 2, we analyzed the posttraining data to evaluate the pattern of adaptation transfer; the posttraining data were normalized relative to the pretraining phase. To facilitate comparisons across syllables and words, we sign-flipped (multiplied by -1) the

³ The formulae of statistical models in this article follow R syntax, where the variable to the left of the tilde ($\sim$) is the dependent variable, and the independent variables are to the right. An interaction of two independent variables is indicated with a colon (e.g., ShiftDirection: Phase). When two independent variables are joined by an asterisk symbol (*), its full expression includes individual variables and all their interaction (e.g., ShiftDirection * Phase = ShiftDirection + Phase + ShiftDirection: Phase). For mixed-effect models, fixed effects do not use parentheses, and random effects are included in parentheses. Within each parenthesis, a random intercept is placed to the right of a pipe (|), and its corresponding random slope is specified to the left.

$F1$ and $F2$ change data as if participants always received / i /-shift in the initial syllable and / æ /-shift in the second syllable. For example, if a participant received the / æ /-shift in the initial syllable, all formant change values from this participant were multiplied by -1 . The derived dependent variables were labeled $\Delta F1_{\text{transfer}}$ and $\Delta F2_{\text{transfer}}$. The independent variable ShiftOrder, which coded whether the initial syllable received / i /-shift or / æ /-shift, was included in all initial formulae to address the potential impact of this sign-flip practice.

We first tested the maintenance of adaptation during posttraining at the group level by analyzing the trained word *bedhead*, starting with the formulae below (two levels of position: first-syllable, second-syllable).

$$\Delta F1_{\text{transfer}} | \Delta F2_{\text{transfer}} \sim \text{Position} * \text{ShiftOrder} + (\text{Position} * \text{ShiftOrder} | \text{Speaker}). \quad (2)$$

In order to evaluate the transfer of learned adaptation to untrained words, we identified which individuals exhibited evidence of adaptation when producing *bedhead* in the posttraining phase. Bootstrap resampling (iterations = 10,000) was applied per participant for each position to obtain 95% confidence intervals (CIs); those without overlap in their 95% CIs in the expected directions were classified as adapters (17 participants in the $F1$ dimension, 11 participants in the $F2$ dimension). Next, we analyzed the $\Delta F1_{\text{transfer}}$ and $\Delta F2_{\text{transfer}}$ of the transfer words from only these adapters, as participants who fail to adapt will, by definition, not show any transfer of learning. First, the trained syllables in the transfer words were compared to their counterparts in the trained word. To avoid rank deficiency, which could negatively influence the reliability of mixed models (Scandola & Tidoni, 2024), a derived variable ProsodicPosition divided the relevant word chunks into four categories: trained (*bedhead* and *bedhead*; coded as the reference), original prosodic position (*bedspread* and *redhead*), isolated monosyllables (*bed* and *head*), and the opposite prosodic position (*deathbed* and *headrest*). The initial formulae are shown below, where the variable “Syllable” distinguishes “bed” and “head.”

$$\Delta F1_{\text{transfer}} | \Delta F2_{\text{transfer}} \sim \text{Syllable} * \text{ProsodicPosition} * \text{ShiftOrder} + (\text{Syllable} * \text{ProsodicPosition} * \text{ShiftOrder} | \text{Speaker}). \quad (3)$$

Separately, the novel syllables in transfer words were compared with their counterparts (based on prosodic position) in the trained word. The initial formulae are shown below, where the variable “Position” differentiates the first-syllable and second-syllable positions, and the variable “Training” separates the trained word from the rest.

$$\Delta F1_{\text{transfer}} | \Delta F2_{\text{transfer}} \sim \text{Position} * \text{Training} * \text{ShiftOrder} + (\text{Position} * \text{Training} * \text{ShiftOrder} | \text{Speaker}). \quad (4)$$

Finally, we checked whether the disyllabic transfer words were produced with the same stress placement as the trained word *bedhead* during the pretraining and posttraining phases by building models that compared the first-syllable and second-syllable

positions, targeting four acoustic cues commonly associated with English lexical stress. The initial formulae were the following. If the transfer words are produced with the same stress placement as the trained word *bedhead*, we should see no main effect of “Word.”

$$f_0 | \text{Duration} | \text{Intensity}_{\text{Mean}} | \text{Intensity}_{\text{Max}} \sim \text{Word} * \text{Position} * \text{Phase} + (\text{Word} * \text{Position} * \text{Phase} | \text{Speaker}). \quad (5)$$

Results

Our hypothesis relies critically on adaptation to opposing auditory perturbations. For this reason, $F1$ and $F2$ were analyzed and reported separately as the projected perturbations on each dimension, as the / i /-shift and / æ /-shift were not necessarily opposing in $F1$ – $F2$ space due to the particularity of each participant’s vowel space. Because the results of the $F1$ and $F2$ analyses are in broad agreement (see the [Supplemental Materials](#) for details), we report statistical results of only the $F1$ analyses, except when the $F1$ and $F2$ analyses diverge.⁴

Experiment 1

Participants adapted separately to the opposing perturbations applied to the two syllables of *bedhead* in both $F1$ and $F2$ in the expected directions: For the / i /-shift (downward $F1$ and upward $F2$), speakers raised their produced $F1$ and lowered their produced $F2$; for the / æ /-shift (upward $F1$ and downward $F2$), speakers lowered their produced $F1$ and raised their produced $F2$ (Panels A and C in [Figure 2](#)). At the individual level, 90% of speakers produced a more positive mean $F1$ change in the / i /-shift than the / æ /-shift in the analyzed hold phase, and 90% also maintained this pattern in the analyzed washout phase ([Figure 2B](#)); all speakers produced a more negative mean $F2$ change in the / i /-shift than the / æ /-shift in both the analyzed hold and washout phases ([Figure 2D](#)). The selected $F1$ model ($R^2 = 0.228$) was the following.

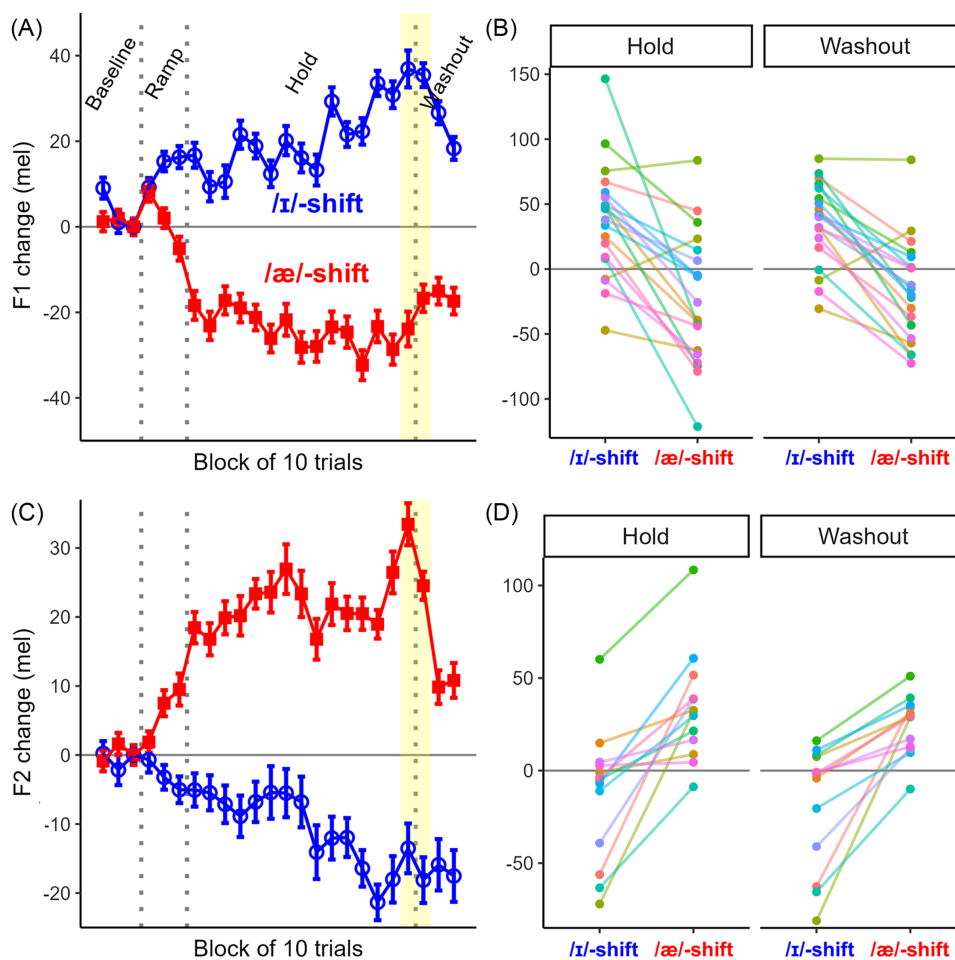
$$\Delta F1 \sim \text{ShiftDirection} + \text{Phase} + (\text{ShiftDirection} + \text{Phase} | \text{Speaker}). \quad (6)$$

As expected, $F1$ change in the / i /-shift was significantly positive (36.17 ± 2.56 mels, $p < .001$), but negative in the / æ /-shift (-20.28 ± 2.59 mels, $p = .04$), yielding a significant difference between the two syllables, a main effect of ShiftDirection: $\chi^2(1) = 40.45$, $R^2 = 0.228$, $d = 1.10$, $p < .001$. No other term in the selected model was significant ($p > .29$), indicating no detectable $F1$ change in the two analyzed windows.

In brief, Experiment 1 confirmed that speakers adapted separately to opposing formant perturbations applied to the two syllables in *bedhead* in the expected directions; this adaptation was maintained immediately after removing the perturbation.

⁴ 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.

Figure 2
Experiment 1 Results



Note. Data points show across-participant means and standard errors (blocks of 10 trials); the shading highlights trials that entered statistical analysis. (A and C) Changes in $F1$ (A) and $F2$ (C) in response to formant perturbations; the dotted vertical lines separate experimental phases. (B and D) Individual speakers' $F1$ (B) and $F2$ (D) changes in the analyzed hold and washout phases; colors represent individual speakers. See the online article for the color version of this figure.

Experiment 2

Experiment 2 further tested the transfer of adaptation to disyllabic words with the same stress placement but with different segmental content; this was achieved by having participants produce the transfer words before the baseline phase (pretraining) and after the hold phase (posttraining).

Training Results

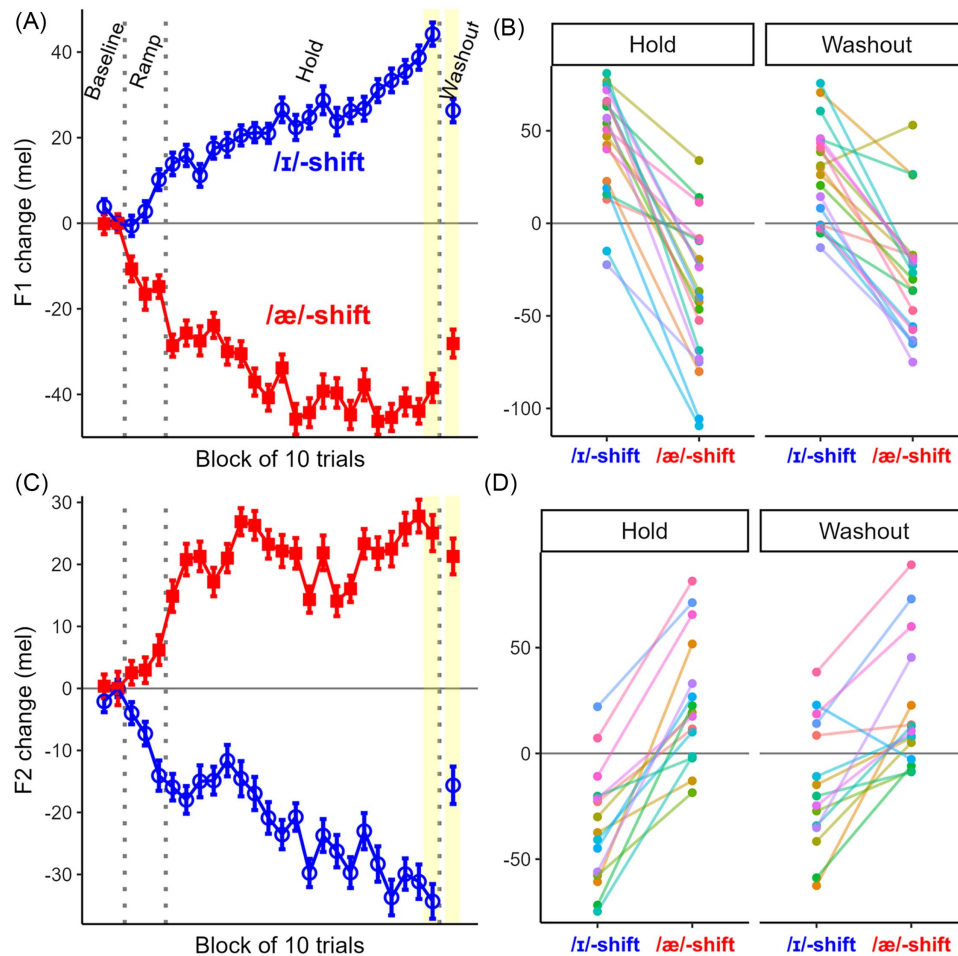
Like Experiment 1, Experiment 2 participants adapted separately to the opposing perturbations applied to two syllables in *bedhead*: For the /ɪ/-shift (downward $F1$ and upward $F2$), speakers raised their produced $F1$ and lowered their produced $F2$; for the /æ/-shift (upward $F1$ and downward $F2$), speakers lowered their produced $F1$ and raised their produced $F2$ (Panels A and C in Figure 3). Unlike Experiment 1, which had no detectable decline of formant changes

in the analyzed washout phase, the washout phase in Experiment 2 showed reduced adaptation, which was expected as the hold and washout phases in Experiment 2 were further separated by 80 posttraining trials. At the individual level, all speakers produced a more positive mean $F1$ change in the /ɪ/-shift than the /æ/-shift in the hold phase, and 95% maintained this pattern in the washout phase (Figure 3B); all speakers produced a more negative mean $F2$ change in the /ɪ/-shift than the /æ/-shift in both the hold and washout phases (Figure 3D). The selected $F1$ model ($R^2 = 0.402$) was the following.

$$\Delta F1 \sim \text{ShiftDirection} + \text{Phase} + \text{ShiftDirection} : \text{Phase} \\ + (\text{ShiftDirection} + \text{Phase} | \text{Speaker}). \quad (7)$$

The $F1$ change in the /ɪ/-shift was significantly positive (35.27 ± 1.99 mels, $p < .001$), but negative in the /æ/-shift (-33.30 ± 2.33 mels, $p < .001$), yielding a significant difference between the two syllables, a main effect of ShiftDirection: $\chi^2(1) = 101.12$, $R^2 = 0.391$, $d = 1.60$,

Figure 3
Experiment 2 Training Results



Note. Data points show across-participant means and standard errors (blocks of 10 trials); the shading covers trials that entered statistical analysis. (A and C) Changes in *F1* (A) and *F2* (C) in response to formant perturbations; the dotted vertical lines separate experimental phases. (B and D) Individual speakers' *F1* (B) and *F2* (D) changes in the analyzed hold and washout phases; colors represent individual speakers. See the online article for the color version of this figure.

$p < .001$. Moreover, the separation between the /ɪ/-shift and the /æ/-shift was greater in the hold phase than in the washout phase, as indicated by a significant interaction of ShiftDirection and Phase, $\chi^2(1) = 43.52$, $R^2 = 0.026$, $p < .001$. Post hoc comparisons also found that the *F1* change in the /ɪ/-shift showed a significant decline from the hold phase to the washout phase ($p = .013$), but no such difference was detected in the /æ/-shift ($p = .14$). The /ɪ/-shift had a significantly more positive *F1* change than the /æ/-shift in both phases (both $p < .001$). The *F1* change differed significantly from zero in both shift directions in both phases (hold: /ɪ/-shift = 44.21 ± 2.73 mels, /æ/-shift = -38.48 ± 3.29 mels; washout: /ɪ/-shift = 26.34 ± 2.75 mels, /æ/-shift = -28.13 ± 3.27 mels; all $p < .001$). No other term in the selected model was significant ($p > .51$).

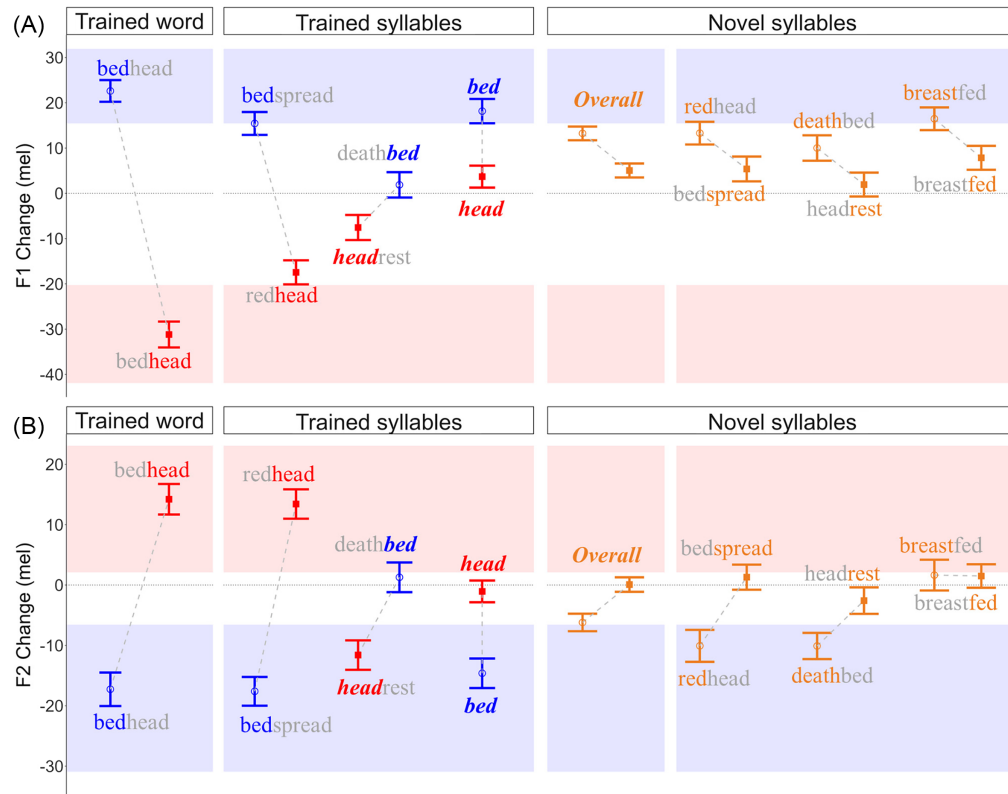
In brief, Experiment 2 replicated the major findings of Experiment 1: Speakers adapted separately to opposing formant perturbations applied to the two syllables in *bedhead* in the expected direction;

this adaptation was maintained but showed signs of a reduction after removing the perturbation.

Transfer Results

To analyze the transfer of learned adaptation, the formant change data in the posttraining phase were normalized relative to the pretraining phase and underwent sign-flip operations as if our participants always received /ɪ/-shift in the initial syllable and /æ/-shift in the second syllable. Figure 4 summarizes the transfer results in the posttraining phase, where the shading shows 95% CIs at the group level based on bootstrap resampling of the trained word (adapters only; iterations = 10,000). The formant changes of the two syllables of *bedhead* were in opposite directions in both *F1* and *F2*, and there was no overlap in the CIs, indicating the maintenance of the learned adaptation under masking noise. Transfer in the trained

Figure 4
Experiment 2 Transfer Results



Note. The figure shows the results of formant changes ($F1$ in A, $F2$ in B) in the posttraining phase based on data from adapters ($/i/$ -shift in the initial syllable; $/æ/$ -shift in the second syllable; data from participants who received the opposite perturbation order were sign-flipped), showing means and standard errors across 10 trials. The shading shows 95% confidence intervals based on bootstrap resampling of the trained word at the group level; only data from adapters were included in analyses of adaptation transfer (17 participants in the $F1$ dimension, 11 participants in the $F2$ dimension). See the online article for the color version of this figure.

syllables was not uniform across words and positions (Figure 4, middle), suggesting that prosody at least plays a modulatory role in adaptation transfer. More importantly, the directionality of formant changes in the novel syllables mirrored the trained syllables (Figure 4, right), although with a reduced magnitude, suggesting adaptation transfer based on prosody alone. See Table 1 for a summary of statistical differences between upward and downward shifts for each condition. Below, we report how the initial and the second syllables, which received opposing perturbations, diverged in light of a potential independent representation of prosodic structure.

The Maintenance of Adaptation. Adaptation was maintained in the trained word *bedhead* during the posttraining phase at the group level: Its $F1$ change was positive in the initial syllable and negative in the second syllable, opposing the $F1$ perturbations (Figure 4, left). The $F1$ model that examined the separation between the two syllables had the following formula ($R^2 = 0.309$).

$$\Delta F1_{\text{transfer}} \sim \text{Position} + (\text{Position}|\text{Speaker}). \quad (8)$$

Consistent with the results from the training phase, $F1$ change was positive in the first-syllable “bed” (20.49 ± 2.33 mels, $p < .001$) and negative in the second-syllable “head” (-27.80 ± 2.87 mels, $p < .001$), with a significant difference between the two, a main effect of position, $\chi^2(1) = 48.55$, $R^2 = 0.309$, $d = 1.32$, $p < .001$. Based on the 95% CIs from the bootstrap resampling, 17 of the 20 participants (85%) individually exhibited adaptation in *bedhead* in this post-training phase (see the Supplemental Materials for details), with significantly more positive $F1$ change in the initial syllable than in the second syllable. Data from only these 17 adapters were analyzed to evaluate the transfer patterns, as reported below.

Comparison of the Trained Word and the Trained Syllables. Adaptation did not transfer equally to all trained syllables (Figure 4, middle): (a) The largest transfer occurred when the trained syllables occurred in the original prosodic position

Table 1

Summary of Adaptation Maintenance (Trained Syllables) and Adaptation Transfer (Novel Syllables), as Reflected by the Separation Between Syllables Receiving /t/-Shift and /æ/-Shift in Experiment 2

Condition	F1 (N = 17) Mean perturbation = 110.00 mels		F2 (N = 11) Mean perturbation = 58.90 mels	
	<i>d</i>	<i>p</i>	<i>d</i>	<i>p</i>
Trained word (<i>bedhead</i>)	1.60	<u><.001</u>	1.16	<u><.001</u>
Trained syllables				
Original (<i>bedspread/redhead</i>)	0.99	<u><.001</u>	1.25	<u><.001</u>
Opposite (<i>deathbed/headrest</i>)	0.26	<u>.023</u>	0.51	<u><.001</u>
Monosyllabic (<i>bed/head</i>)	0.43	<u>.001</u>	0.61	<u><.001</u>
Novel syllables				
Overall	0.24	<u>.062</u>	0.26	<u><.001</u>
With trained syllable in original position (<i>bedspread, redhead</i>)	0.23	.25	0.46	<u><.001</u>
With trained syllable in opposite position (<i>deathbed/headrest</i>)	0.23	.25	0.33	<u>.01</u>
Two novel syllables (<i>breastfed</i>)	0.26	.25	0.01	.96

Note. Conditions with significant differences between the two shift directions are shown underlined and in italics; marginally significant separations are underlined.

(*bedspread* and *redhead*); (b) when the trained syllables occurred in isolated monosyllables, transfer was complete in *bed* but absent in *head*; (c) when the trained syllables occurred in the opposite prosodic position (*deathbed* and *headrest*), they tended to follow the direction of the trained syllable, though with a substantially reduced magnitude. The selected *F1* model was the following ($R^2 = 0.223$).

$$\begin{aligned} \Delta F1_{\text{transfer}} \sim & \text{Syllable} + \text{ProsodicPosition} \\ & + \text{Syllable} : \text{ProsodicPosition} + \text{ShiftOrder} \\ & + \text{ProsodicPosition} : \text{ShiftOrder} \\ & + (\text{Syllable}|\text{Speaker}). \end{aligned} \quad (9)$$

The “bed” syllables had significantly larger *F1* changes than the “head” syllables, indicated by a significant main effect of syllable, $\chi^2(1) = 157.28$, $R^2 = 0.135$, $d = 0.77$, $p < .001$. Whether the trained syllables stayed in their original prosodic position also modulated the *F1* change significantly, a main effect of ProsodicPosition, $\chi^2(3) = 66.86$, $R^2 = 0.02$, $p < .001$. The selected model further contained a significant interaction of the prosodic position of the trained syllables and the counterbalanced order of the shifts, ProsodicPosition:ShiftOrder: $\chi^2(3) = 26.95$, $R^2 < 0.001$, $p < .001$, yet no difference was found based on whether participants first received the /æ/-shift or /t/-shift, ShiftOrder: $\chi^2(1) = 0.018$, $R^2 = 0.001$, $p = .67$, including comparisons across all prosodic positions (all $p > .08$), suggesting that the influence of the order of the shifts on transfer was limited.

Importantly, the magnitude of the separation between the trained syllables “bed” and “head” differed depending on their prosodic position in the untrained words, an interaction of Syllable and ProsodicPosition, $\chi^2(3) = 119.37$, $R^2 = 0.104$, $p < .001$. Post hoc comparisons found that the “bed” syllables had significantly more positive *F1* change than the “head” syllables, consistent with the transfer of learned adaptation across all prosodic positions (all $p < .024$). The separation of “bed” (23.82 ± 2.39 mels, significantly positive, $p < .001$) and “head” (-30.78 ± 2.86 mels, significantly

negative, $p < .001$) was the largest in the trained word *bedhead* ($d = 1.61$, $p < .001$), which was also significantly larger than the other prosodic positions, indicating incomplete adaptation transfer ($d = 0.99$, 0.43 , and 0.26 ; all $p < .001$; details below).

When trained syllables occurred in their original prosodic position, transfer was present but incomplete. This was true for both *bedspread* (15.59 ± 2.53 mels, significantly positive, $p = .009$), which showed a significantly smaller change than *bedhead* ($d = 0.23$, $p = .034$), and *redhead* (-17.26 ± 2.66 mels, significantly negative, $p < .001$), which showed a significantly smaller change than *bedhead* ($d = 0.39$, $p < .001$). A direct comparison of *bedspread* and *redhead* showed that they differed significantly from each other ($d = 0.99$, $p < .001$).

In monosyllables, transfer was essentially complete in the monosyllabic *bed* (18.09 ± 2.68 mels, significantly positive, $p = .002$), with no difference from the reference *bedhead* ($d = 0.13$, $p = .15$). Conversely, transfer was entirely absent in the monosyllabic *head*, whose *F1* change did not differ from zero (3.61 ± 2.43 mels, $p = .86$) and was, in fact, numerically in the opposite direction as *bedhead*; it also differed significantly from *bedhead* ($d = 1.02$, $p < .001$). A direct comparison of the change in *F1* in these monosyllables showed a significant difference between the two ($p = .002$, $d = 0.43$).

Finally, transfer was small, but present, for trained syllables in the opposite prosodic position. In *deathbed*, the *F1* change was in the same direction as *bedhead* (2.75 ± 2.80 mels, no difference from zero, $p = .86$), but of significantly smaller magnitude ($d = 0.62$, $p < .001$); in *headrest*, the *F1* change was in the same direction as *bedhead* (-7.02 ± 2.77 mels, no difference from zero, $p = .38$), but of significantly smaller magnitude ($d = 0.66$, $p < .001$). Crucially, a direct comparison of the *F1* change in *deathbed* and *headrest* showed a significant difference ($d = 0.26$, $p = .024$). In other words, although the change of prosodic position modulated adaptation transfer, the direction of adaptation transfer followed the direction of change in the trained syllables (same segmental content) rather than the position in the prosodic frame.

The results for changes in *F2* largely mirrored those for *F1*, with the exception of when the trained syllables occurred in the opposite

prosodic position: While $F1$ changes in this case patterned with the trained segmental content (e.g., “bed” in both *bedhead* and *deathbed* had positive $F1$ change), the $F2$ changes patterned instead with prosodic position. In *deathbed* (0.36 ± 2.46 mels, no difference from zero, $p = .99$), the $F2$ change was in the same direction as the second syllable in *bedhead* (12.70 ± 2.46 mels, marginally positive, $p = .094$), but of significantly smaller magnitude ($d = 0.50$, $p < .001$). In *headrest* (-11.28 ± 2.44 mels, no difference from zero, $p = .11$), the $F2$ change was in the same direction as the first syllable in *bedhead* (-18.42 ± 2.77 mels, significantly positive, $p = .039$), with no significant difference in magnitude ($d = 0.21$, $p = .13$).

In summary, adaptation did not transfer equally to all trained syllables. Transfer was the largest when a trained syllable remained in its original prosodic position (also paired with a novel syllable), although even in this case, transfer was incomplete. In the opposite prosodic position, there was a difference in the expected direction between “bed” in *deathbed* and “head” in *headrest*, suggesting some transfer of learning; however, when examining the syllables individually against zero or the reference *bedhead*, significant results were detected, suggesting small transfer. Finally, the first-syllable position also seemed more conducive to adaptation transfer than the second-syllable position, as the transfer was complete in the isolated *bed* but was absent in the isolated *head*. However, because all disyllabic stimuli have the same lexical stress placement (primary stress on the first syllable and secondary stress on the second syllable), it is unclear whether this effect is due to syllable position per se or to lexical stress placement. Regardless, and taking the full pattern of results into consideration, these findings indicate that transfer was influenced by both segmental and prosodic structure.

Comparison of the Trained Word and the Novel Syllables.

Examining adaptation transfer in novel syllables provides the most direct test of the independence of prosody, as this transfer can only occur if prosodic structure retains some independent representation after its coordination with segmental content. Consistent with such an independent representation, novel syllables showed formant changes according to their position in the prosodic frame: Syllables in the first-syllable position (primary stress) showed larger $F1$ changes than those in the second-syllable position (secondary stress), mirroring the pattern observed in the trained syllables (Figure 4, right); see Supplemental Figure S4 for averages of individual participants. While the overall pattern was similar for both $F1$ and $F2$, statistical analyses of the two formants yielded slightly different results. The $F1$ model for comparing the trained word *bedhead* and the novel syllables was the following ($R^2 = 0.188$).

$$\begin{aligned} \Delta F1_{\text{transfer}} \sim & \text{Position} + \text{Training} + \text{Position} : \text{Training} \\ & + \text{ShiftOrder} + \text{Position} : \text{ShiftOrder} \\ & + \text{Training} : \text{ShiftOrder} \\ & + \text{Position} : \text{Training} : \text{ShiftOrder} + (\text{Position} \\ & + \text{Training} | \text{Speaker}). \end{aligned} \quad (10)$$

The $F1$ change in the first-syllable position was significantly positive (18.54 ± 1.29 mels, $p < .001$), and the second-syllable position was significantly negative (-12.69 ± 1.49 mels, $p = .011$),

yielding a significant difference between the two, a main effect of position: $\chi^2(1) = 35.33$, $R^2 = 0.093$, $d = 0.54$, $p < .001$. The trained and novel syllables also differed significantly, a main effect of training, $\chi^2(1) = 13.05$, $R^2 = 0.026$, $p < .001$.

Importantly, the separation between the two positions was larger in the trained syllables than in the novel syllables, indicating reduced transfer in the novel syllables, a significant interaction of position and training, $\chi^2(1) = 61.98$, $R^2 = 0.028$, $p < .001$. Critically, post hoc comparisons revealed the presence of incomplete transfer in the novel syllables based on their prosodic positions. When the novel syllables were examined according to their position in the prosodic frame, the $F1$ change in the first-syllable position (13.39 ± 1.51 mels, significantly positive, $p = .021$) was in the same direction, but reduced, compared to *bedhead* ($d = 0.28$, $p = .010$), indicating incomplete transfer; in the second-syllable position, there was no direct evidence for transfer, as the $F1$ change in the novel syllables (5.43 ± 1.54 mels, no difference from zero, $p = .30$) was in the opposite direction as the reference *bedhead* and differed significantly from it ($d = 0.28$, $p = .010$). Directly comparing the first- and second-syllable positions showed a significant difference between the syllables in the trained word ($d = 1.60$, $p < .001$); the difference in the novel syllables was numerically in the same direction, but this difference was not significant ($d = 0.24$, $p = .062$). The $F1$ model also detected a significant interaction of position, training, and ShiftOrder, $\chi^2(1) = 7.34$, $R^2 = 0.003$, $p = .007$ (see Supplemental Table S1), suggesting that the specific order of the applied shifts modulated the $F1$ transfer results; however, even after accounting for order effects, the main differences between syllable positions and trained/novel syllables were still found. No other term in the selected model was significant ($p > .085$).

The $F2$ model for comparing the trained word *bedhead* and the novel syllables was the following ($R^2 = 0.096$).

$$\begin{aligned} \Delta F2_{\text{transfer}} \sim & \text{Position} + \text{Training} + \text{Position} : \text{Training} \\ & + (\text{Position} + \text{Training} | \text{Speaker}). \end{aligned} \quad (11)$$

The $F2$ change in the first-syllable position was significantly negative (-12.70 ± 1.31 mels, $p = .02$), but the change in the second-syllable position, while in the expected direction, did not differ from zero (6.18 ± 1.14 mels, $p = .21$). Nonetheless, there was a significant difference between the two syllable positions, a main effect of position: $\chi^2(1) = 127.32$, $R^2 = 0.093$, $d = 0.49$, $p < .001$. The separation between the two positions was larger in the trained syllables than in the novel syllables, indicating reduced transfer in the novel syllables, a significant interaction of position and training, $\chi^2(1) = 56.73$, $R^2 = 0.044$, $p < .001$.

Similar to $F1$ results, post hoc comparisons revealed the presence of incomplete transfer in the novel syllables based on their prosodic positions. When individual novel syllables were examined, the $F2$ change in the first-syllable position (-6.29 ± 1.51 mels, no difference from zero, $p = .15$) was in the same direction, but reduced, compared to *bedhead* ($d = 0.42$, $p = .030$); the $F2$ change in the second-syllable position did not differ from zero (-0.02 ± 1.54 mels, $p = .99$) but differed significantly from the reference *bedhead* ($d = 0.63$, $p = .030$). Directly comparing the first- and second-syllable positions showed significant differences between the syllables in

both trained words ($d = 1.16$, $p < .001$) and, crucially, in the novel syllables ($d = 0.26$, $p < .001$).

Because the presence of a trained syllable in a transfer word (e.g., *bedspread*) may influence adaptation transfer across positions (cf. the transfer from *bé* to both syllables of *bépé* in Caudrelier et al., 2018), we also modeled the trained and the novel syllables by making finer distinctions in the novel syllables. The results broadly agreed with the above reports (see the Supplemental Materials for details).

To summarize, adaptation transferred from the trained word *bedhead* to novel syllables with the same prosodic frame based on their prosodic position within that frame. Although the magnitude of this transfer was relatively small compared to segmental-based transfer in trained syllables, these results nonetheless provide evidence for an independent representation of prosody separate from the segmental content of speech: In the novel syllables, the separation between the first-syllable position (primary stress) and the second-syllable position (secondary stress) mirrored the separation of “bed” and “head” in *bedhead* in both $F1$ and $F2$. Examination of the two syllable positions separately showed that transfer was present but incomplete in the first-syllable position, but absent in the second-syllable position, mirroring the primacy of the first-syllable/primary-stress position observed when the trained syllables in *bedhead* were produced as monosyllables.

Prosodic Consistency Between the Trained and Transfer Words

We checked acoustic cues commonly associated with prosody (median f_0 , mean RMS intensity, maximum RMS intensity, and duration) and found that the transfer words were produced with similar prosodic realizations as the trained word *bedhead* (Figure 5; see the Supplemental Materials for detailed statistical results). As expected for words with first-syllable primary stress, the first-syllable position was consistently produced with higher median f_0 than the second-syllable position across words and phases (all $p < .001$; Figure 5A); the same pattern (higher f_0 in the first-syllable position) held when only examining adapters in the $F1$ dimension ($n = 17$; all $p < .002$) and in the $F2$ dimension ($n = 11$; all $p < .001$). Similarly, the first syllables were produced with higher intensity (both mean RMS intensity and maximal RMS intensity, all $p \leq 0.05$).

Conversely, the first-syllable position consistently had a shorter duration than the second-syllable position across phases and words (all $p < .001$); the same pattern held when examining only adapters in the $F1$ dimension ($n = 17$; all $p < .001$) and in the $F2$ dimension ($n = 11$; all $p < .001$). However, the longer duration in the second-syllable position likely reflects phrase-final lengthening, that is, the temporal expansion of the final syllable in an utterance. Phrase-final lengthening is seemingly universal: It is consistently reported in English (e.g., Oller, 1973; Wightman et al., 1992), as well as cross-linguistically (e.g., Seifart et al., 2021). Critically, phrase-final lengthening has been shown to result in overall longer durations of a final unstressed syllable compared to initial stressed syllables in phrase-final trochees like *bedhead* (Berkovits, 1994), strongly suggesting that the durational differences found in our data do not reflect the presence of primary stress on the final syllable.

To summarize, patterns of produced f_0 and intensity clearly indicate that the primary stress fell on the first syllable. Moreover, across all the acoustic cues, the transfer words were produced with

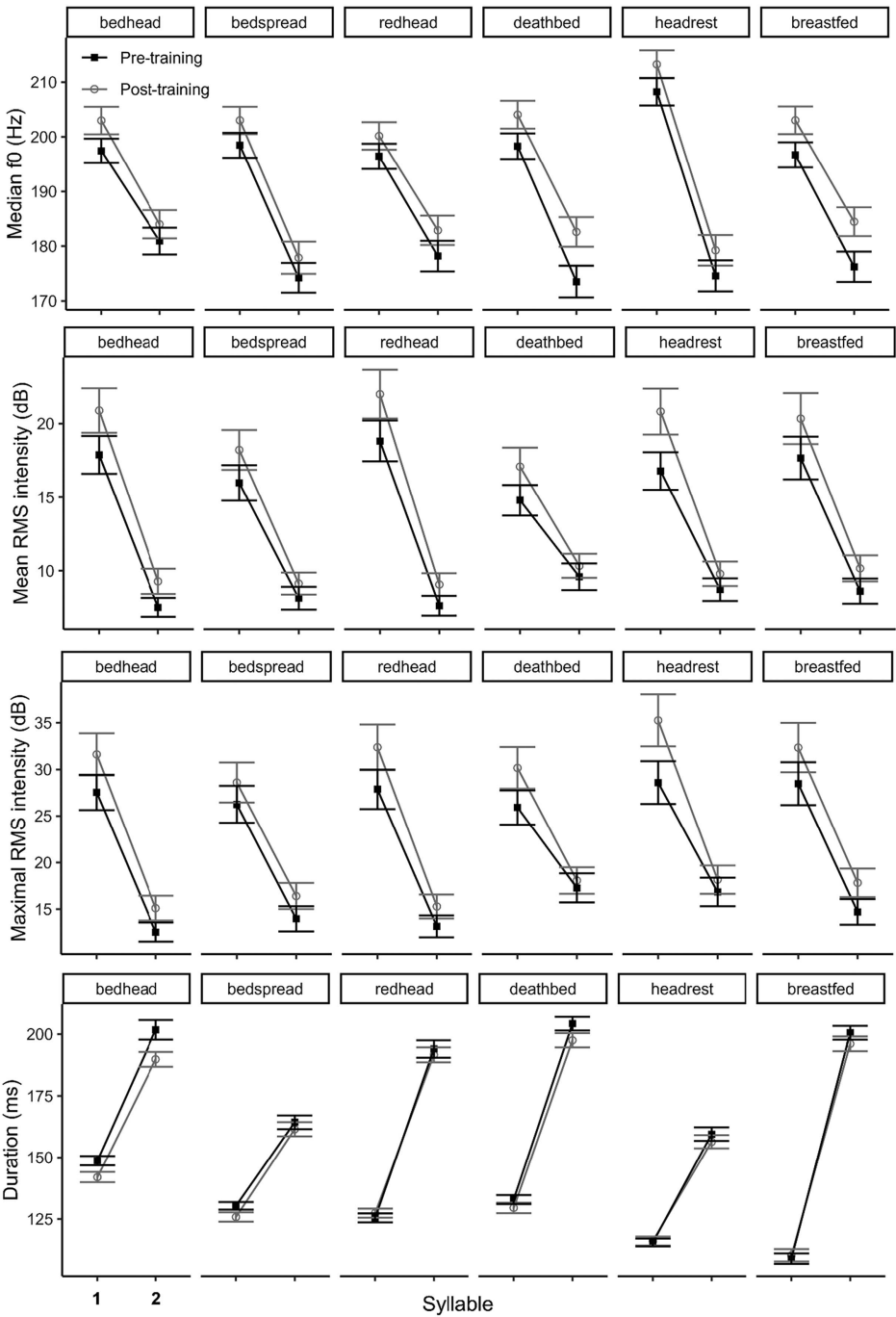
the same prosodic structure as the trained word *bedhead*, corroborating the claim that the shared prosodic structure is the primary driver of the adaptation transfer in novel syllables.

Discussion

In running speech, segments and prosody co-occur in time, so there must be a mechanism that coordinates their production during planning. Previous research suggests that segments still have an independent representation after this segment-prosody coordination, but it remains unclear whether prosody also has an independent representation. Here, we addressed this question using a novel paradigm: evaluating the transfer pattern of speech sensorimotor adaptation from a trained word to untrained words containing both trained and novel syllables, all containing the same vowel /e/. Since we applied opposing formant perturbations to the first and second syllables of the trained word, adaptation purely driven by the vowel /e/ is expected to be zero. Speakers adapted separately to the opposing formant perturbations applied to the two syllables in the trained word *bedhead*, and this learned adaptation transferred both to the same syllables in new words as well as to novel syllables. In syllables that occurred in the trained word, transfer was largest when trained syllables remained in their original prosodic position (e.g., *bedspread*), but was reduced (or absent) when the trained syllables occurred in the opposite prosodic position (e.g., *deathbed*). Transfer was small but consistently present in novel syllables (e.g., *breastfed*) according to their position in the prosodic frame. Across both trained and novel syllables, the first-syllable position with primary stress appeared more conducive to adaptation transfer than the second-syllable position with secondary stress. The patterns of transfer observed in the trained syllables are largely compatible with both major classes of models concerning the coordination of segmental content and prosodic structure. The prosody-first account (e.g., Keating & Shattuck-Hufnagel, 2002; Turk & Shattuck-Hufnagel, 2020) proposes that prosody is planned before segments, with the latter filling in the slots in a prosodic frame created by the former. The segment-first account (e.g., Levelt, 1989; Levelt et al., 1999), in contrast, posits that segments are planned first, followed by prosody being superimposed on already planned segments. Regardless of the timing order of the segments and prosody, both accounts suggest that segmental structure retains some independent representation apart from its prosodic context, so both predict that the transfer of adaptation will be the largest when trained syllables remain in their original prosodic position but reduced in magnitude (yet still present) when trained syllables change their prosodic positions. Both predictions were borne out in our study: Transfer from *bedhead* was larger in *bedspread* and *redhead* ($d = 0.99/1.25$ in $F1/F2$) than in *deathbed* and *headrest* ($d = 0.43/0.51$ in $F1/F2$).

However, these two models differ in their predictions about adaptation transfer based on the prosodic frame alone. In our data, we found small but consistent adaptation transfer in the novel syllables based on their prosodic positions. Crucially, as the adaptation transfer in the novel syllables was most likely due to the shared prosodic structure, our study presents the first behavioral evidence supporting the presence of an independent prosodic representation in motor plans. Such an interpretation is more consistent with the prosody-first account (e.g., Keating & Shattuck-Hufnagel, 2002; Turk & Shattuck-Hufnagel, 2020), which assumes

Figure 5
Experiment 2 Measures of Lexical Stress



Note. The figure shows the prosodic cues of disyllabic words in Experiment 2 during the pretraining and posttraining phases, showing the consistent prosodic realizations between the trained word and other transfer words. All Experiment 2 participants were included; qualitatively similar results were found when only adapters identified in the *F1* and *F2* dimensions were tested ($n = 17$ and 11 , respectively). RMS = root mean square.

the independence of prosody in speech production. Our results are at odds with the segment-first account (e.g., [Levelt, 1989](#); [Levelt et al., 1999](#)), which predicts no independent prosodic representation after its integration/coordination with segments. In the existing literature,

there is abundant evidence for the independence of prosody at the phrase and sentence levels (see [Shattuck-Hufnagel, 2015, 2020](#), for overviews). Our results show that such an independent prosodic representation is also present at the word level. Nevertheless, our

results cannot answer whether word-level prosodic structure is planned before or after segmental content or if segmental and prosodic structures are planned simultaneously (cf. Zeng et al., 2025a).

Notably, our results in favor of the independent representation of word-level prosodic structure contradict previous findings that failed to elicit prosodic priming of spoken words. It may be that priming, in general, is simply ineffective in tapping into prosodic representations, as suggested by the difficulty of eliciting priming based on lexical tones in Mandarin Chinese, a core phonological distinction in that language (Zeng, 2022). Indeed, the failure of priming to elicit prosodic effects is not restricted to speech production; perceptual experiments involving priming have also failed to uncover prosodic effects (Cutler, 1986; Sereno & Lee, 2015). For example, Sereno and Lee (2015), who used a primed auditory lexical decision task, essentially replicated the results of Myers and Watson (2019) in spoken word recognition: The priming effect, measured as gain in reaction time relative to unrelated controls, was the largest when overlap occurred in lexical tone and segments (rù-rù; diacritics show lexical tone categories), weaker when there was only segmental overlap (rǔ-rù), and absent when there was only lexical tone overlap (shà-rù). However, later perceptual studies using more sensitive measures, such as eye-tracking, indeed found evidence for the continued use of lexical tone in speech perception in both native and nonnative listeners (Qin et al., 2019), corroborating the idea that priming is ineffective at detecting prosodic effects. More broadly, procedures relying on reaction time and accuracy, even without a priming component, are overall unsuccessful in finding prosodic effects in speech perception (Cutler & Chen, 1997; Cutler & Clifton, 1984; Slowiaczek, 1990; Wiener & Turnbull, 2016) and speech production (Myers & Watson, 2021). Accordingly, the lack of prosodic effects reported in previous studies of speech production (Myers & Watson, 2019, 2021; Roelofs & Meyer, 1998) should be reinterpreted as a lack of evidence for an independent prosodic representation instead of evidence of its absence. An important open question worthy of further investigation is why prosodic effects are more difficult to detect in procedures such as priming.

We found that adaptation transfer based on position within the prosodic frame was substantially smaller than transfer based on shared segmental content (e.g., the nearly complete transfer from the first syllable in *bedhead* to the isolated word *bed*), suggesting that the segmental representation may be stronger than the prosodic representation. However, and in contrast with previous results from priming studies, we observed several cases where adaptation transfer reflected primarily prosodic structure rather than segmental content. For example, transfer is absent from the second syllable (with secondary stress) in *bedhead* to the isolated word *head* (with primary stress). More directly, in our *F2* results, when trained syllables occurred in a novel prosodic position, the transfer of adaptation followed the trained syllable position rather than the trained segmental content (negative *F2* changes in *bedspread* and *headrest* and positive *F2* changes in *redhead* and *deathbed*), indicating that transfer in this case is primarily determined by prosodic structure. Conversely, our *F1* results are more consistent with the suggestion from previous studies (i.e., a stronger segmental representation that is modulated, but not dominated, by prosody): Transfer in the trained syllables largely followed trained segments, as identical syllables had the same direction of formant change regardless of their prosodic position

(positive *F1* change in *bedspread* and *deathbed* and negative *F1* change in *redhead* and *headrest*).

The disparity in *F1* and *F2* results may be explained by weaker segmental transfer in *F2* (difference between /l/-shift and /æ/-shift; 95% CI [6.56, 14.55] mels) than in *F1* (95% CI [14.51, 23.21] mels), which may have allowed prosodic structure to surface as the primary driver of adaptation and the subsequent transfer in the *F2* dimension. We see two potential reasons why segment-based learning and transfer may have been reduced in *F2* compared to *F1*. First, the magnitude of segmental transfer may vary with the magnitude of the auditory perturbation. In the present study, the magnitude of the *F2* perturbations ($M = 56.40$ mels) was considerably smaller than the *F1* perturbations ($M = 107.62$ mels). As the magnitude of adaptation generally scales with the magnitude of the perturbation (e.g., Katseff et al., 2012; MacDonald et al., 2010), smaller perturbations may have led to reduced segmental-based learning and transfer in the *F2* dimension. In contrast, the prosodic transfer in novel syllables was of similar size in *F1* and *F2* ($d = 0.25$, $p = .062$ in *F1*; $d = 0.26$, $p = .001$), potentially suggesting that prosody-based learning is relatively independent of perturbation magnitude. Alternatively, it may be that *F2* provides a weaker cue to vowel categories than *F1* (at least among the perturbed /e/ vowel compared to the neighboring /l/ and /æ/ vowels toward which the auditory feedback was perturbed). Given that segmental-based articulatory changes in the motor system in response to auditory errors are generally larger when those errors cross a category boundary (see C. A. Nizioletk & Guenther, 2013, for online compensation across vowel boundaries and Bourguignon et al., 2014, for adaptation across lexical boundaries), the stronger categorical changes induced by *F1* perturbations may have resulted in more segment-based learning in this dimension, allowing prosodic effects to emerge in *F2*. Notably, this should be the case even if the magnitude of the perturbations in the two dimensions had been equal. Follow-up experiments applying the same perturbation magnitude to *F1* and *F2* could resolve this question.

While the present study provides evidence for the independence of prosodic structure, it is unclear exactly what structural component of prosody this adaptation transfer reflects. If we consider that the primary aspect of the metrical frame is the placement of lexical stress, our results show that primary stress is more conducive to adaptation transfer than secondary stress (see Bakst & Nizioletk, 2021, for similar results). However, there are several other possibilities. First, it is possible that the transfer observed here is based on primary stress, rather than a metrical frame per se. This would be consistent with the nearly complete transfer observed from “bed” in “bedhead” to the monosyllable “bed,” where “bed” receives primary stress in both cases, compared to the lack of transfer to the monosyllable “head,” where there is a mismatch in stress. This account could also explain the results from the transfer to novel syllables, where transfer was observed primarily in the stressed syllables. Conversely, these same effects would also be consistent with prosodic transfer relying on syllable position in the word, with primacy given to the first syllable. From an information theory perspective, sounds occurring early in a word are more informative than later ones (e.g., one does not need to hear all sounds in “cucumber” before it emerges as the only viable word). The importance of early sounds is also universally embraced by models of spoken word recognition (e.g., Marslen-Wilson & Tyler, 1980; McClelland & Elman, 1986; Norris & McQueen, 2008). In speech production, the initial segments and syllables may also enjoy a

similar special status, which could be conducive to adaptation transfer. Follow-up experiments that test adaptation transfer to disyllabic words with the primary stress on the second-syllable position could clarify which feature(s) of prosody (metrical frame, lexical stress, syllable position) drive the prosodic transfer in novel syllables.

In this study, we focus on word-level prosody, and our results show for the first time that speakers are able to adapt to distinct auditory perturbations of different syllables within the same prosodic word. While previous work has shown that speakers can adapt to such opposing perturbations when they are applied to separate monosyllabic words sharing the same vowel (i.e., *bed* and *head*; Rochet-Capellan & Ostry, 2011), our results showed that even when united in a single word, separate syllables sharing the same vowel can show differential adaptation (e.g., “bedhead” → “bidhad”). The transfer of learned adaptation based on a shared disyllabic word-level prosodic structure in Experiment 2 also clearly indicates that, in the current experiments, *bedhead* was produced as a single unified word rather than as separate, isolated syllables. If “bedhead” was produced as two isolated monosyllabic words, we would expect to see (near)-complete transfer of adaptation to the isolated monosyllables “bed” (confirmed) and “head” (disconfirmed); it could be that isolated monosyllabic words are treated like the first syllable in disyllabic words. Regardless, that prosody enables and modulates adaptation transfer indicates that the disyllabic words are produced as a single unified word, instead of a simple concatenation of isolated monosyllables. The ability of speakers to adapt to different auditory perturbations within a single word is also consistent with a recent study from our lab that reports strong evidence for multisyllabic motor planning units, likely corresponding to words (Zeng et al., 2025a).

Finally, although our results provide evidence supporting an independent prosodic representation at the word level in speech production, there are several unanswered questions regarding the generalizability of this conclusion. First, as discussed above, future research is needed to clarify which aspect(s) of prosody (the metrical frame, lexical stress, syllable position) enable the transfer of adaptation—answering this question is critical for understanding what aspect(s) of prosodic structure have an independent representation in speech production. Second, our study only examined transfer to words sharing the same vowel /e/ that occurred in the trained word. While this was critical here to ensure that any segment-based learning canceled out, an independent prosodic representation should additionally transfer to words sharing a prosodic frame but with different vowels (e.g., *backpack*, *picnic*). We believe this is highly likely, given the established transfer of adaptation across vowels within monosyllables (Rochet-Capellan et al., 2012), but it remains to be tested. Third, we examined only a single (trochaic) metric frame. However, especially given the consistency of the transfer results, there is no strong reason to assume that our results cannot be generalized to other metrical frames (e.g., iambic rather than trochaic feet) in English, as well as in other languages with similar prosodic structures. However, establishing the universality of our results will require investigation into prosodic features not present in English (e.g., Chinese lexical tone, Japanese pitch accent); findings from such studies will bring important insights regarding the independence of prosodic representation across languages.

Conclusion

Using speech sensorimotor adaptation, we confirmed that speakers could adapt to opposing formant perturbations applied to the two syllables of *bedhead*. The learned adaptation showed the largest transfer when the trained syllables remained in their original prosodic position (e.g., *bedspread*), indicating that transfer was enabled by segments and prosody. Crucially, adaptation transferred partially to disyllabic words with the same prosodic structure but novel syllables (e.g., *breastfed*), suggesting the presence of an independent representation of a word-level prosodic frame. By revealing straightforward behavioral evidence for an independent word-level prosodic representation in speech production for the first time, our findings have important implications for models of speech production, especially regarding the relationship between segmental and prosodic planning. At the minimum, prosody should not be considered secondary to segments, as has been claimed based on previous studies that have failed to elicit priming based only on prosodic structure. More generally, this study showcases how speech sensorimotor adaptation can provide a highly sensitive behavioral measure that may be useful for uncovering effects in psycholinguistic planning that are difficult to detect with other more standard procedures.

References

- Bakst, S., & Niziolet, C. A. (2021). Effects of syllable stress in adaptation to altered auditory feedback in vowels. *The Journal of the Acoustical Society of America*, 149(1), 708–719. <https://doi.org/10.1121/10.0003052>
- 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* (Version 1.1-34) [R]. <https://CRAN.R-project.org/package=lme4>
- Berkovits, R. (1994). Durational effects in final lengthening, gapping, and contrastive stress. *Language and Speech*, 37(3), 237–250. <https://doi.org/10.1177/002383099403700302>
- 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/iaul/. *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
- 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>
- Cutler, A. (1986). Forbear is a homophone: Lexical prosody does not constrain lexical access. *Language and Speech*, 29(3), 201–220. <https://doi.org/10.1177/002383098602900302>

- Cutler, A., & Chen, H.-C. (1997). Lexical tone in Cantonese spoken-word processing. *Perception & Psychophysics*, 59(2), 165–179. <https://doi.org/10.3758/BF03211886>
- Cutler, A., & Clifton, C., Jr. (1984). The use of prosodic information in word recognition. In H. Bouma & D. G. Bouwhuis (Eds.), *Attention and performance X: Control of language processes* (pp. 183–196). Erlbaum. <https://hdl.handle.net/11858/00-001M-0000-0013-35C9-4>
- 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>
- Gussenhoven, C., & Chen, A. (Eds.). (2021). *The Oxford handbook of language prosody*. Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780198832232.001.0001>
- Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics, Theory and Applications*, 6(2), 65–70. <http://www.jstor.org/stable/4615733>
- Katseff, S., Houde, J., & Johnson, K. (2012). Partial compensation for altered auditory feedback: A tradeoff with somatosensory feedback? *Language and Speech*, 55(2), 295–308. <https://doi.org/10.1177/0023830911417802>
- Keating, P., & Shattuck-Hufnagel, S. (2002). A prosodic view of word form encoding for speech production. *UCLA working papers in phonetics* (pp. 112–156). Department of Linguistics.
- 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
- Kuznetsova, A., Brockhoff, P. B., Christensen, R. H. B., & Jensen, S. P. (2020). *lmerTest: Tests in linear mixed effects models* (Version 3.1-3) [R]. <https://CRAN.R-project.org/package=lmerTest>
- Lametti, D. R., Smith, H. J., Watkins, K. E., & Shiller, D. M. (2018). Robust sensorimotor learning during variable sentence-level speech. *Current Biology*, 28(19), 3106–3113.e2. <https://doi.org/10.1016/j.cub.2018.07.030>
- Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. The MIT Press. <https://doi.org/10.7551/mitpress/6393.001.0001>
- 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](https://doi.org/10.1016/0010-0277(92)90038-J)
- Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22(1), 1–38. <https://doi.org/10.1017/S0140525X99001776>
- Lieberman, P. (1960). Some acoustic correlates of word stress in American English. *The Journal of the Acoustical Society of America*, 32(4), 451–454. <https://doi.org/10.1121/1.1908095>
- MacDonald, E. N., Goldberg, R., & Munhall, K. G. (2010). Compensations in response to real-time formant perturbations of different magnitudes. *The Journal of the Acoustical Society of America*, 127(2), 1059–1068. <https://doi.org/10.1121/1.3278606>
- Marslen-Wilson, W., & Tyler, L. K. (1980). The temporal structure of spoken language understanding. *Cognition*, 8(1), 1–71. [https://doi.org/10.1016/0010-0277\(80\)90015-3](https://doi.org/10.1016/0010-0277(80)90015-3)
- 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>
- McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., & Sonderegger, M. (2017). Montreal forced aligner: Trainable text-speech alignment using Kaldi. *Interspeech*, 498–502, 498–502. <https://doi.org/10.21437/Interspeech.2017-1386>
- McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18(1), 1–86. [https://doi.org/10.1016/0010-0285\(86\)90015-0](https://doi.org/10.1016/0010-0285(86)90015-0)
- Myers, B. R., & Watson, D. G. (2019). Paying the meter: Effect of metrical similarity on word lengthening. *Psychonomic Bulletin & Review*, 26(6), 1941–1947. <https://doi.org/10.3758/s13423-019-01635-4>
- Myers, B. R., & Watson, D. G. (2021). Evidence of absence: Abstract metrical structure in speech planning. *Cognitive Science*, 45(8), Article e13017. <https://doi.org/10.1111/cogs.13017>
- 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. (2014). When bed goes bad: How the brain can fix mistakes in speech while they happen. *Frontiers for Young Minds*, 2(2). <https://doi.org/10.3389/frym.2014.00002>
- Niziolek, C. A., & Guenther, F. H. (2013). Vowel category boundaries enhance cortical and behavioral responses to speech feedback alterations. *Journal of Neuroscience*, 33(29), 12090–12098. <https://doi.org/10.1523/JNEUROSCI.1008-13.2013>
- Niziolek, C. A., & Houde, J. F. (2015). *wave_viewer: First release* [Computer software]. Zenodo. <https://doi.org/10.5281/zenodo.13839>
- Norris, D., & McQueen, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review*, 115(2), 357–395. <https://doi.org/10.1037/0033-295X.115.2.357>
- Oh, Y. M., Coupé, C., Marsico, E., & Pellegrino, F. (2015). Bridging phonological system and lexicon: Insights from a corpus study of functional load. *Journal of Phonetics*, 53, 153–176. <https://doi.org/10.1016/j.wocn.2015.08.003>
- Oller, D. K. (1973). The effect of position in utterance on speech segment duration in English. *The Journal of the Acoustical Society of America*, 54(5), 1235–1247. <https://doi.org/10.1121/1.1914393>
- Plag, I., Kunter, G., & Schramm, M. (2011). Acoustic correlates of primary and secondary stress in North American English. *Journal of Phonetics*, 39(3), 362–374. <https://doi.org/10.1016/j.wocn.2011.03.004>
- Qin, Z., Tremblay, A., & Zhang, J. (2019). Influence of within-category tonal information in the recognition of Mandarin-Chinese words by native and non-native listeners: An eye-tracking study. *Journal of Phonetics*, 73, 144–157. <https://doi.org/10.1016/j.wocn.2019.01.002>
- 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/>
- 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>
- Rochet-Capellan, A., Richer, L., & Ostry, D. J. (2012). Nonhomogeneous transfer reveals specificity in speech motor learning. *Journal of Neurophysiology*, 107(6), 1711–1717. <https://doi.org/10.1152/jn.00773.2011>
- Roelofs, A., & Meyer, A. S. (1998). Metrical structure in planning the production of spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24(4), 922–939. <https://doi.org/10.1037/0278-7393.24.4.922>
- Scandola, M., & Tidoni, E. (2024). Reliability and feasibility of linear mixed models in fully crossed experimental designs. *Advances in Methods and Practices in Psychological Science*, 7(1), 1–21. <https://doi.org/10.1177/25152459231214454>
- Seifart, F., Strunk, J., Danielsen, S., Hartmann, I., Pakendorf, B., Wichmann, S., Witzlack-Makarevich, A., Himmelmann, N. P., & Bickel, B. (2021). The extent and degree of utterance-final word lengthening in spontaneous speech from 10 languages. *Linguistics Vanguard*, 7(1), 20190063. <https://doi.org/10.1515/lingvan-2019-0063>
- Sereno, J. A., & Lee, H. (2015). The contribution of segmental and tonal information in Mandarin spoken word processing. *Language and Speech*, 58(2), 131–151. <https://doi.org/10.1177/0023830914522956>
- Shattuck-Hufnagel, S. (1992). The role of word structure in segmental serial ordering. *Cognition*, 42(1–3), 213–259. [https://doi.org/10.1016/0010-0277\(92\)90044-1](https://doi.org/10.1016/0010-0277(92)90044-1)

- Shattuck-Hufnagel, S. (2015). Prosodic frames in speech production. In M. A. Redford (Ed.), *The handbook of speech production* (pp. 419–444). Wiley. <https://doi.org/10.1002/9781118584156.ch19>
- Shattuck-Hufnagel, S. (2020). The role of phrase-level prosody in speech production planning. In C. Gussenhoven & A. Chen (Eds.), *The Oxford handbook of language prosody* (pp. 522–538). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780198832232.013.34>
- Slowiaczek, L. M. (1990). Effects of lexical stress in auditory word recognition. *Language and Speech*, 33(1), 47–68. <https://doi.org/10.1177/002383099003300104>
- Speer, S., & Blodgett, A. (2006). Prosody. In M. J. Traxler & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (pp. 505–537). Academic Press. <https://doi.org/10.1016/B978-012369374-7/50014-6>
- Stevens, S. S., Volkman, J., & Newman, E. B. (1937). A scale for the measurement of the psychological magnitude pitch. *The Journal of the Acoustical Society of America*, 8(3), 185–190. <https://doi.org/10.1121/1.1915893>
- Surendran, D., & Levow, G.-A. (2004). The functional load of tone in Mandarin is as high as that of vowels. *Speech Prosody*, 99–102, 99–102. <https://doi.org/10.21437/SpeechProsody.2004-23>
- Tang, D. L. (2024). Using altered auditory feedback to study pitch compensation and adaptation in tonal language speakers. *Frontiers in Human Neuroscience*, 18, Article 1364803. <https://doi.org/10.3389/fnhum.2024.1364803>
- 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>
- Turk, A., & Shattuck-Hufnagel, S. (2020). *Speech timing: Implications for theories of phonology, speech production, and speech motor control*. Oxford University Press. <https://doi.org/10.1093/oso/9780198795421.001.0001>
- 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>
- Wiener, S., & Turnbull, R. (2016). Constraints of tones, vowels and consonants on lexical selection in Mandarin Chinese. *Language and Speech*, 59(1), 59–82. <https://doi.org/10.1177/0023830915578000>
- Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M., & Price, P. J. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *The Journal of the Acoustical Society of America*, 91(3), 1707–1717. <https://doi.org/10.1121/1.402450>
- Zeng, Y. (2022). *Spoken word production of mandarin monosyllabic words: From lexical selection to form encoding* [Doctoral dissertation]. University of Kansas. <https://www.proquest.com/docview/2714408640/abstract/1C89A83164194E67PQ/1>
- Zeng, Y., Niziolek, C. A., & Parrell, B. (2025a). Simultaneous acquisition of multiple auditory–motor transformations reveals suprasyllabic motor planning in speech production. *Journal of Experimental Psychology: General*, 154(6), 1681–1698. <https://doi.org/10.1037/xge0001744>
- Zeng, Y., Niziolek, C. A., & Parrell, B. (2025b, May 19). *Transfer of sensorimotor adaptation reveals independent prosodic representation after segment-prosody coordination in speech production*. <https://osf.io/d4t6s>

Received January 13, 2025

Revision received August 28, 2025

Accepted September 12, 2025 ■