

1 Audiomotor prediction errors drive speech
2 adaptation even in the absence of overt movement

3

4 Benjamin Parrell^{1,2}, Chris Naber¹, Olivia A. Kim³, Caroline A. Nizolek^{1,2}, & Samuel D.
5 McDougale^{4,5}

6

7 ¹ Waisman Center, University of Wisconsin–Madison

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

9 ³ Program in Neuroscience, Bates College

10 ⁴ Department of Psychology, Yale University

11 ⁵ Wu Tsai Institute, Yale University

12

13

14 Abstract

15 Observed outcomes of our movements sometimes differ from our expectations. These sensory
16 prediction errors recalibrate the brain's internal models for motor control, reflected in alterations
17 to subsequent movements that counteract these errors (motor adaptation). While leading theories
18 suggest that all forms of motor adaptation are driven by learning from sensory prediction errors,
19 dominant models of speech adaptation argue that adaptation results from integrating time-
20 advanced copies of corrective feedback commands into feedforward motor programs. Here, we
21 tested these competing theories of speech adaptation by inducing planned, but not executed,
22 speech. Human speakers (male and female) were prompted to speak a word and, on a subset of
23 trials, were rapidly cued to withhold the prompted speech. On standard trials, speakers were
24 exposed to real-time playback of their own speech with an auditory perturbation of the first formant
25 to induce single-trial speech adaptation. Speakers experienced a similar sensory error on
26 movement cancelation trials, hearing a perturbation applied to a recording of their speech from a
27 previous trial at the time they would have spoken. Speakers adapted to auditory prediction errors
28 in both contexts, altering the spectral content of spoken vowels to counteract formant
29 perturbations even when no actual movement coincided with the perturbed feedback. These
30 results build upon recent findings in reaching, and suggest that prediction errors, rather than
31 corrective motor commands, drive adaptation in speech.

32 Introduction

33 Implicit motor adaptation is a prime example of the brain's predictive capacities. In the standard
34 framework, motor adaptation occurs when the predicted sensory consequences of a descending
35 motor command are compared against feedback, and an internal model of those motor
36 commands is adjusted to reduce observed discrepancies (i.e., sensory prediction errors) (Wolpert
37 and Kawato, 1998; Wolpert and Flanagan, 2016; Krakauer et al., 2019). One lingering question
38 in this framework is what serves as the input signal for the generation of sensory predictions that
39 drive motor learning. In some models, that input signal is an efference copy generated from the
40 descending motor command, which is then used to predict future states (Scott, 2004; Shadmehr
41 and Krakauer, 2008; Houde and Nagarajan, 2011; Parrell et al., 2019). In other work, the input
42 signal is defined as a higher-level motor plan (Day et al., 2016; Sheahan et al., 2016, 2018;
43 McDougle et al., 2017; Vyas et al., 2020). It remains an open question whether motor planning or
44 execution are sufficient for the computation of sensory predictions and errors, and whether there
45 are common requirements for adaptation across motor subsystems.

46 In the case of speech adaptation, the prevailing idea is that ongoing feedback during actual
47 execution of speech commands drives adaptation to audiomotor prediction errors (Tourville and
48 Guenther, 2011; Guenther, 2016). Conversely, a recent model suggests that adaptation relies
49 instead on auditory errors generated by the comparison of auditory reafference with auditory
50 predictions generated by a task or planning-level controller rather than by motor efference (Kim
51 et al., 2023a, 2023b). To date, it has been difficult to distinguish between these competing theories
52 as they make similar predictions in existing speech adaptation paradigms. That said, several lines
53 of evidence from upper-limb adaptation support the notion that an “upstream” motor plan is
54 sufficient to drive adaptation. One piece of evidence for this is the discovery that internal signals
55 preceding a movement, such as planning “lead-in” and “follow-through” motions, can partition
56 otherwise catastrophically interfering motor memories (Sheahan et al., 2016, 2018). Indeed, a
57 range of similar contextual pre-cueing effects have been observed in motor adaptation, supporting
58 the idea that preparatory signals seed learning computations within the adaptation system
59 (Wolpert et al., 2011; Howard et al., 2013; Heald et al., 2018, 2021; Vyas et al., 2020; Avraham
60 et al., 2022; Churchland and Shenoy, 2024).

61 We have recently developed an experimental paradigm in limb control that is capable of teasing
62 apart plan- vs. motor-based prediction error in sensorimotor adaptation, which we used to
63 establish that implicit upper-limb motor adaptation could be induced in response to sensory errors
64 that were unaccompanied by movements (Kim et al., 2022). Human subjects performed a
65 “Go/NoGo” task that involved planning, but occasionally not executing, reaches to visual targets.
66 On certain critical trials, subjects successfully inhibited reach commands but still observed a
67 “virtual” sensory prediction error delivered via a visual cursor rotated relative to the planned
68 reaching direction. These trials induced adaptation in subsequent movements, even though no
69 movement had accompanied the error. In our view, these results offer particularly compelling
70 support for the aforementioned plan-based model of adaptation.

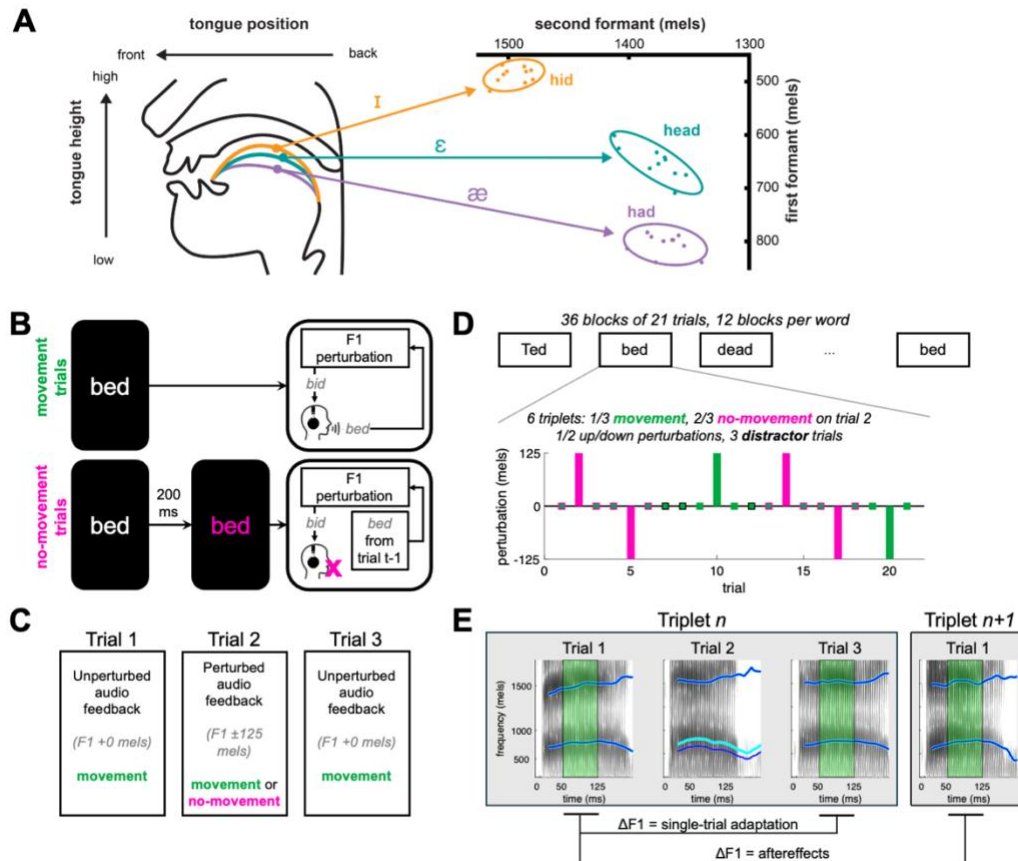
71 Here we applied this design to speech adaptation in order to answer one of the major outstanding
72 questions about sensorimotor adaptation in speech: whether ongoing feedback from executed

73 speech movements is necessary for adaptation or whether speech adaptation can leverage
74 sensory predictions based only on planned speech acts. Moreover, by expanding this paradigm
75 from a laboratory reaching task to speech, an entirely novel motor domain that differs in several
76 critical ways from limb control (bulbar vs. spinal control, hydrostatic control of the tongue vs.
77 control of limb joint angles, auditory vs. visual feedback, etc), this work provides a critical test of
78 the domain-generalty of our previous findings. Similar results in speech and reaching would
79 strongly support a shared neurocomputational mechanism driving sensorimotor adaptation.

80 Results

81 We implemented a task that allowed us to measure speech adaptation both with and without
82 movement execution during perceived auditory prediction errors (Figure 1). Trials were organized
83 into three-trial “triplets”, where the middle trial received a perturbation to the first vowel formant
84 (F1), one of the resonances of the vocal tract determined by the position of the tongue, lips, and
85 jaw which distinguish vowels from one another (Figure 1A). On “movement” trials, participants
86 produced speech when cued by a visual stimulus. To measure adaptation in the absence of
87 movement execution, we had participants occasionally inhibit a cued speech movement while still
88 presenting them with a well-timed audiomotor prediction error (“no-movement” trials), using a
89 formant perturbation applied to a recording of the speech produced on the immediately preceding
90 trial. We measured adaptation for following both movement and no-movement trials as the change
91 in F1 from the trial preceding the perturbed trial to the two trials immediately following the
92 perturbed trial (adaptation and aftereffects, Figure 1E).

93



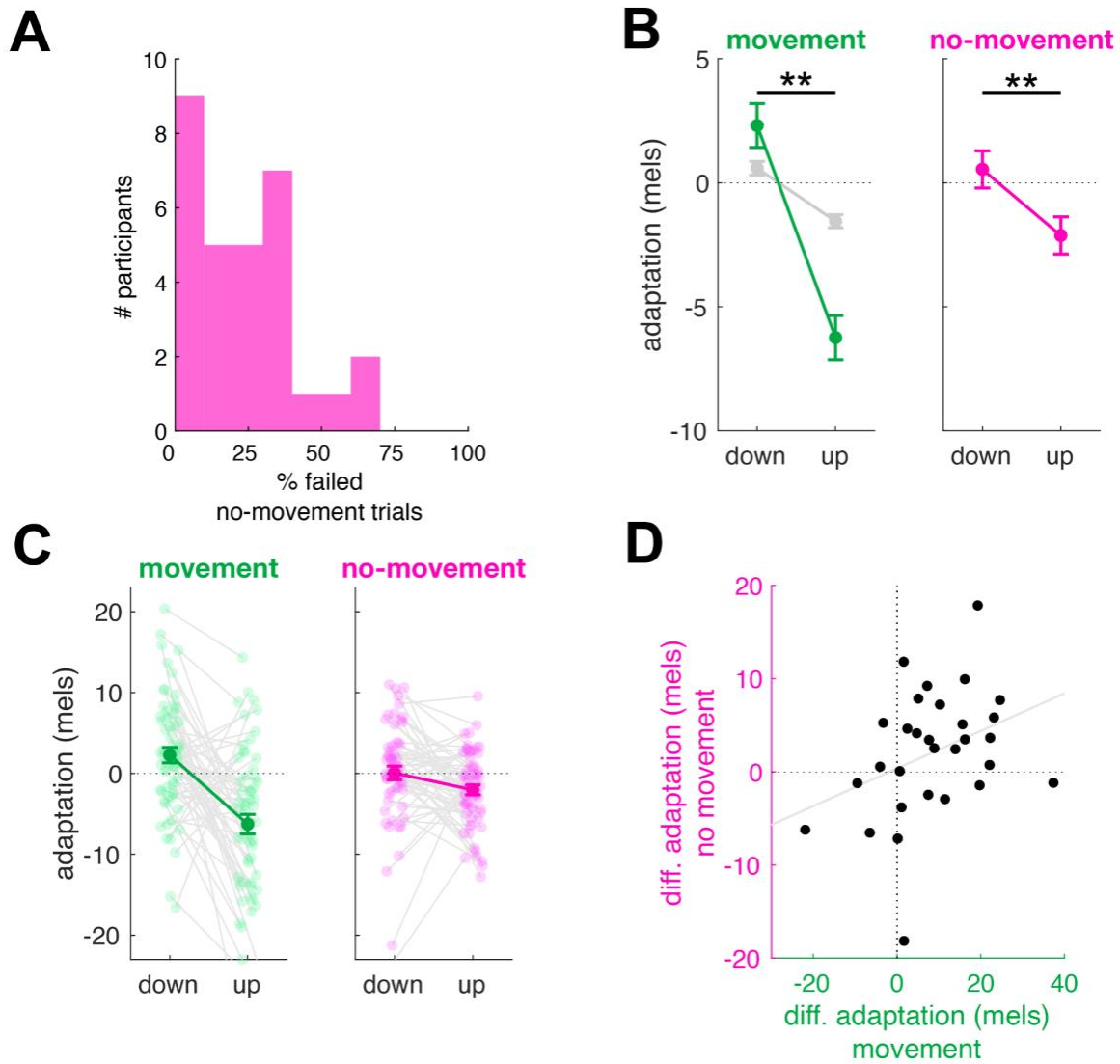
94 *Figure 1: Experimental design. A: Illustration of the relationship between tongue position and*
 95 *vowel formants. The first two vowel formants (F1, F2) serve to distinguish between different*
 96 *vowels. B: Schematic of movement and no-movement trials. In movement trials, a stimulus word*
 97 *appeared on the screen and participants read the word out loud. Participants heard their own*
 98 *voice, with the first vowel formant (F1) perturbed or unperturbed. In no-movement trials, the*
 99 *stimulus word turned red 200 ms after appearing. Participants were instructed not to produce*
 100 *speech on these trials. Instead, the audio recording from the previous trial was played back to the*
 101 *participants with F1 perturbed. C: Triplet design. In each triplet, trials 1 and 3 were always*
 102 *movement trials with no perturbation applied. Trial 2 was variable – a movement or no-movement*
 103 *trial – and always had a perturbation applied. D: Experiment structure. The experiment consisted*
 104 *of 36 blocks which alternated between three stimulus words. Each block consisted of 6 triplets (2*
 105 *with movement middle trials, 4 with no-movement middle trials) and 3 distractors (always*
 106 *movement trials) to disrupt the rhythm of the perturbations. E: Data analysis. Each trial shows a*
 107 *spectrogram of an example speech trial; cyan lines show produced vowel formants (F1 and F2)*
 108 *and blue lines show formants in headphone signal. For each trial, the produced F1 (the lowest*
 109 *frequency formant) was averaged between 50-125 ms after vowel onset (green shaded portion)*
 110 *to avoid coarticulatory effects of the initial consonant in each stimulus word and potential online*
 111 *compensatory effects. Single-trial adaptation was measured as the change in F1 from trial 1 to*
 112 *trial 3 within each triplet. Aftereffects were measured as the change from trial 1 of one triplet to*
 113 *the trial immediately following trial 3 of that triplet (either a distractor or trial 1 of the next triplet).*

114 Despite the instruction to inhibit speech, participants nonetheless produced some acoustic
115 speech material on a subset of no-movement trials (average of 22%, range 3-62%, Figure 2D).
116 The prevalence of these failures suggests that our method was successful in driving participants
117 to plan speech at the onset of the stimulus in these no-movement trials. In order to isolate learning
118 in the absence of movement, only no-movement trials where participants successfully inhibited
119 speech were used to measure adaptation.

120 Because the expected effect size of single-trial adaptation (~3 mels in Hantzsch et al., 2022, Fig
121 2A) is much smaller than the expected standard deviation of F1 in typical speech (~10-15 mels),
122 our primary outcome measurement combined measurements taken from the third trial of each
123 triplet immediately following the perturbation (adaptation) as well as the subsequent trial
124 (aftereffects), in order to increase the power of our analysis and reduce noise. Participants
125 produced robust single-trial learning following perturbations to both movement and no-movement
126 trials (Figure 2A,B, main effect of perturbation direction $F(1,10724) = 59.25$, $p < 1 \times 10^{-13}$, partial R^2
127 = 0.005). For movement trials, F1 following trials with a downward F1 perturbation showed a
128 significantly higher F1 (2.3 ± 0.8 mels) than following trials with an upward F1 perturbation (-6.2
129 ± 0.8 mels, $t(10722) = 7.5$, $p < .0001$, $d = 0.23$). Single-trial adaptation on movement trials was 3-
130 4 larger than in our previous work (Hantzsch et al., 2022), likely reflecting better control of stimuli
131 in the current paradigm (see *Discussion*).

132 Crucially, following no-movement trials, F1 was significantly higher in trials following downward
133 perturbations (0.5 ± 0.7 mels) compared to upward perturbations (-2.1 ± 0.8 mels, $t(10736) = 2.9$,
134 $p = 0.004$, $d = 0.07$). This result supported our hypothesis that speech adaptation could proceed
135 without actual movement execution. The magnitude of adaptation was smaller in no-movement
136 trials compared to movement trials, reflected by a significant interaction between movement
137 condition and perturbation direction ($F(1,10725) = 16.3414$, $p < 0.0001$, partial $R^2 = 0.002$).
138 Overall, the magnitude of adaptation in no-movement trials was roughly one-third of that found in
139 movement trials, similar to our previous results in adaptation of upper limb movements (Kim et
140 al., 2022). (Potential reasons for this replicated drop in effect size are addressed in the
141 *Discussion*.)

142 We then looked at a between-subjects correlation between adaptation effect sizes on movement
143 versus no-movement triplets. Indeed, there was a moderate, though statistically marginal, positive
144 relationship between learning in movement and non-movement trials across participants in the
145 expected direction (Figure 2C, $r = 0.36$, $p = 0.053$).

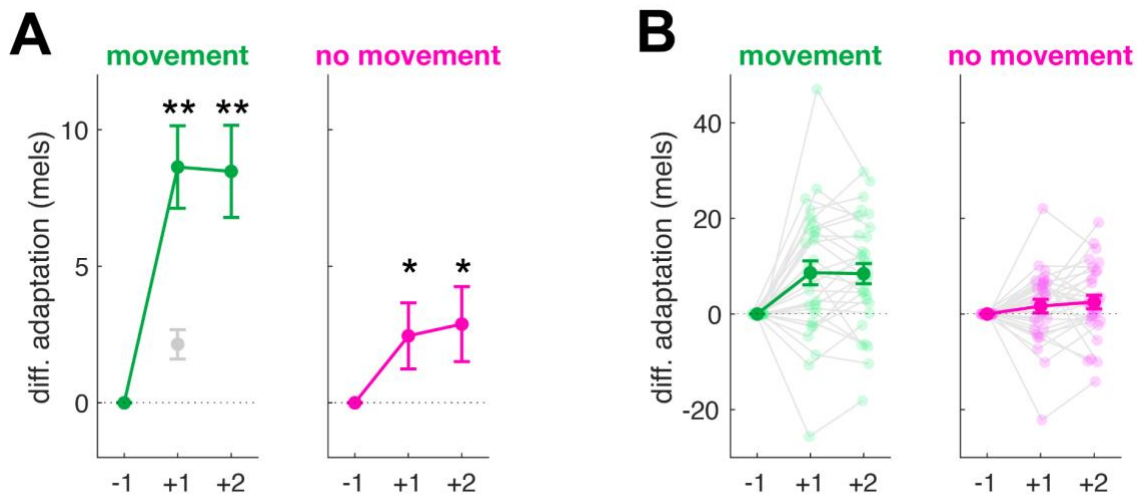


146

147 *Figure 2: Experimental results for the adaptation index. A: The percentage of no-movement trials*
148 *on which participants failed to inhibit vocal production. The average failure rate was 22% across*
149 *participants. These trials were excluded from further analysis. B: F1 adaptation index (see*
150 *Methods) following trials with downward and upward perturbations applied in movement (left) and*
151 *no-movement (right) conditions. Estimated marginal means and standard errors shown in green*
152 *and magenta. Gray indicates single-trial learning observed in Hatzsch et al. (2022). C: As for (B),*
153 *showing data for individual participants. Means and standard error across participants are shown*
154 *in dark, solid colors. Individual participant means shown as lighter colored dots connected by gray*
155 *lines. D: Relationship across participants between the magnitude of differential adaptation*
156 *(responses following downward perturbation minus responses following upward perturbations) in*
157 *movement and non-movement conditions.*

158 As an additional analysis, we considered adaptation (in trials immediately following the
159 perturbation, $t+1$) and aftereffects (in the next following trial, $t+2$) separately (Figure 3). The results
160 for both adaptation and aftereffects were consistent with the combined results. For adaptation,
161 there was a significant main effect of perturbation direction ($F(1,5362.3) = 32.7, p < 1 \times 10^{-7}$, partial
162 $R^2 = 0.006$) as well as a significant interaction between perturbation direction and movement
163 condition ($F(1,5362.3) = 10.2, p = 0.001$, partial $R^2 = 0.002$). The difference in F1 between upward
164 and downward trials was significant for both movement (8.6 ± 1.5 mels, $t(5363) = 5.7, p < 0.0001$,
165 $d = 0.24$) and no-movement conditions (2.5 ± 1.2 mels, $t(5369) = 2.014, p = 0.04, d = 0.07$), though
166 the magnitude of this difference was again smaller for the no-movement condition.

167 The results for aftereffects were similar: Upward and downward perturbations led to differences
168 in F1 in both movement (8.5 ± 1.7 mels, $t(5328) = 4.997, p < .0001, d = 0.22$) and no-movement
169 (2.9 ± 1.4 mels, $t(5341) = 2.110, p = 0.03, d = 0.07$) conditions (main effect of perturbation direction:
170 $F(1,5355) = 27.2, p < 1 \times 10^{-6}$, partial $R^2 = 0.005$); and the magnitude of this difference was smaller
171 in the no-movement condition (interaction: $F(1,5355) = 6.6, p = 0.01$, partial $R^2 = 0.001$).



172

173 *Figure 3: Experimental results separated by trial number. A: Differential adaptation shown*
174 *separately for trials immediately following the perturbation (+1, adaptation) and the next trial (+2,*
175 *aftereffects) in the movement (left) and no-movement (right) conditions. Estimated marginal*
176 *means and standard errors shown in green and magenta. Gray indicates single-trial learning*
177 *observed in Hatzsch et al. (2022). B: As for (A), showing data for individual participants. Means*
178 *and standard error across participants are shown in dark, solid colors while individual participant*
179 *means are shown as lighter colored dots connected by gray lines. * $p < 0.05$; ** $p < 0.005$.*

180 Discussion

181 Here we demonstrated that speech adaptation can occur even when audiomotor errors are not
182 accompanied by movement execution. First, we replicated the result that when participants were
183 exposed to a perturbation of their first vowel formant on a single trial, changes in speech articulator

184 movements that opposed that perturbation were visible on the two trials following the perturbation.
185 These results confirm that single-trial speech adaptation observed in previous paradigms
186 designed to examine online compensatory movements (Hantzsch et al., 2022) can be reliably
187 elicited with our triplet design. Remarkably, adaptation occurred not only when participants heard
188 an auditory perturbation during speech, but also when participants planned to produce speech,
189 but withheld overt speech movement and instead heard playback of their own speech from a
190 previous trial with the perturbation applied. That is, when participants putatively generated a
191 movement plan and a concomitant auditory prediction, they adapted their movements to correct
192 for observed audiomotor prediction errors even though they did not produce any overt speech
193 movements in conjunction with those errors. These results extend, in a new motor and sensory
194 domain, our recent findings in upper-limb visuomotor adaptation (Kim et al., 2022).

195 The current results suggest that audiomotor speech adaptation is driven, at least in part, by using
196 sensory prediction errors to directly update internal models controlling speech articulators. Such
197 updates are consistent with dominant models of sensorimotor learning in limb and oculomotor
198 control (Wolpert et al., 1998; Smith et al., 2006; Shadmehr and Krakauer, 2008; Shadmehr et al.,
199 2010; Hadjiosif et al., 2021). However, these results are inconsistent with the dominant model of
200 sensorimotor adaptation in speech production, which suggests that adaptation is driven by the
201 incorporation of corrective movements generated by a feedback controller in response to sensory
202 errors into future feedforward motor programs (Tourville and Guenther, 2011; Guenther, 2016;
203 Kearney et al., 2020), which itself is partially based on other ideas in limb control (Kawato et al.,
204 1987; Albert and Shadmehr, 2016).

205 Notably, the magnitude of the change observed on the movement triplets (~8 mel difference
206 between upward and downward perturbations) was roughly 2-3 times larger than we previously
207 observed using the same analysis window (Hantzsch et al., 2022). Why did we see this large
208 increase in effect size? One possibility is that our new design, which uses blocked repetition of
209 the same stimuli, induces stronger learning signals than when stimuli are mixed across trials as
210 in previous studies. This could be because adaptation in speech only partially generalizes to
211 untrained words (Rochet-Capellan et al., 2012; Caudrelier et al., 2018), similar to the local spatial
212 generalization of learning observed in reaching (Gandolfo et al., 1996; Krakauer et al., 2000;
213 Donchin et al., 2003). Importantly, the larger effect size we observed in our movement trials may
214 allow for more precise assays of factors that affect sensorimotor adaptation in speech, given that
215 around 100 observations of learning can be obtained in the time it would typically take to obtain
216 a single observation (e.g., mean asymptotic learning) using a more traditional paradigm with
217 extended exposure to a repeated auditory perturbation.

218 Intriguingly, the size of the reduction we observed in speech adaptation between movement and
219 no-movement trials (roughly one-third, Figures 2 and 3) was very similar to the reduction
220 previously observed in a similar reaching task (Kim et al., 2022), pointing to a domain-general
221 explanation. We see several possible explanations for this observed reduction: First, it has been
222 established, both in speech and reaching tasks, that temporal delays between movement and
223 sensory feedback substantially impair adaptation to errors (Kitazawa et al., 1995; Brudner et al.,
224 2016; Schween and Hegele, 2017; Zhou et al., 2017). In speech, the magnitude of adaptation is

225 reduced by roughly 50% when auditory feedback is delayed by only 100ms, and adaptation is
226 essentially absent when delays reach 250-500ms (Max and Maffett, 2015; Shiller et al., 2020).
227 This suggests that the critical comparison driving auditory error processing is highly temporally-
228 specific, particularly in speech. In our paradigm it is impossible to know exactly when speech
229 would have occurred on the no-movement trials; as an estimate, the latency of the speech
230 feedback on a given no-movement trial was matched to the latency on a previous movement trial.
231 This method is likely to have introduced variance between the timing of anticipated and perceived
232 auditory feedback on no-movement trials, potentially leading to a reduction in the magnitude of
233 adaptation.

234 Second, reductions in somatosensory feedback inherent in no-movement trials may have
235 suppressed the degree of adaptation possible. It has been suggested that, in limb control,
236 adaptation may result not from a drive to reduce visual sensory error but rather from the
237 recalibration of somatosensory and other sensory signals (Tsay et al., 2022). Consistent with this
238 idea, putatively disruptive noninvasive stimulation to primary somatosensory cortex following
239 visuomotor reach adaptation substantially decreases how much of this learning is retained
240 (Ebrahimi and Ostry, 2024). Thus, when errors occur in the absence of somatosensory feedback,
241 such as in our no-movement trials, it is possible that adaptation magnitude would be reduced.

242 Finally, it may also be the case that adaptation in speech can be driven by auditory prediction
243 errors both directly through updates to predictive internal models and indirectly through
244 incorporation of previous feedback-based commands into feedforward motor programs. In fact,
245 there is support for this “dual input” idea in tasks requiring adapting to dynamic force-field
246 perturbations during reaching (Albert and Shadmehr, 2016). Thus, an important note here is that
247 while many results point to a plan-based sensory prediction error model of adaptation, both
248 planning and ongoing feedback commands could both provide inputs that generate sensory
249 predictions. Future work can directly assay if and how both sources of information fuel adaptation.

250 An additional putative learning mechanism that may be connected to the current study is
251 observational learning. It is possible that additional contributions from observation-based learning
252 could be relevant here, as subjects in our task observed errors without moving (Mattar and
253 Gribble, 2005; Pawlowsky et al., 2023). However, we believe it is unlikely that observational
254 learning fully explains our results: In speech production, producing a word after an auditory
255 presentation of the same word typically drives speakers to change their production to be *closer*
256 to the auditory stimulus, a process known as “phonetic convergence” (Goldinger, 1998; Shockley
257 et al., 2004; Mattar and Gribble, 2005; Babel, 2012; Babel and Bulatov, 2012; Pawlowsky et al.,
258 2023).

259 Although direct neurophysiological data related to speech adaptation is limited, our results are
260 consistent with invasive and non-invasive imaging studies that have examined sensory prediction
261 in speech. When we speak, auditory cortical activity is suppressed relative to when we passively
262 listen to the same sounds, a process thought to be driven by the cancellation of auditory
263 reafference using predictive internal models (Curio et al., 2000; Houde et al., 2002; Ventura et al.,
264 2009; Flinker et al., 2010). This suppression effect is reduced in less prototypical productions,
265 suggesting that the prediction is based on a plan (or target) rather than motor efference (Niziolek

266 et al., 2013; Tang et al., 2023; Beach et al., 2024). Recent work has shown that suppression is
267 modulated during adaptation, and that the degree of modulation predicts the degree of learning,
268 strongly suggesting that these predictions (and the resulting sensory prediction errors) play a
269 critical role in adaptation (Kim et al., 2023b). Single-unit recordings in marmosets indicate that
270 suppression of auditory cortex firing begins hundreds of milliseconds ahead of vocalization
271 initiation (Eliades and Wang, 2003), and stimulus-specific cortical modulation has been
272 demonstrated in this pre-vocalization window in humans (Daliri and Max, 2016), suggesting that
273 vocalization planning alone can modulate auditory cortex activity and may be sufficient to support
274 error computation.

275 While our results cannot directly answer whether such sensory errors drive learning through
276 updates to forward models or control policies, they are nonetheless consistent with the general
277 consensus that adaptation is driven by a cerebellar-dependent process based on sensory
278 prediction errors (Bastian, 2006; Flanagan et al., 2003; Haith & Krakauer, 2013; Miall & Wolpert,
279 1996; Shadmehr et al., 2010; Wolpert & Kawato, 1998). Our results, along with our previous
280 results in upper-limb adaptation, strongly suggest that these predictions are seeded, at least in
281 part, by a higher-level planning process.

282 There are two caveats about our no-movement condition. First, although single-trial adaptation
283 was robust in the no-movement condition, the overall effect was relatively small (~2 mels, $d =$
284 0.07). Though this magnitude is roughly equivalent to our previous work demonstrating single-
285 trial adaptation (in movement trials) when stimulus words varied across trials (Hantzsch et al.,
286 2022)), speech, in general, is substantially more variable than this, with a standard deviation in
287 F1 of roughly 10-15 mels. Our triplet design allowed for a high number of observations per
288 participant in order to look for this “needle in the haystack.” Second, while we excluded no-
289 movement trials with any overt sound production of any kind, it is possible that participants
290 nonetheless produced some subtle muscular activity on a subset of these trials. However, there
291 is some evidence, though from a slightly different task, that articulatory movements and voicing
292 frequently co-occur when stop signals occur at a similar latency after the go signal as in the current
293 study (Tilsen and Goldstein, 2012). This suggests that the acoustic signal used as a criteria for
294 exclusion here is likely to also exclude trials with overt movement – as such, we do not believe
295 that potential latent muscular activity substantially changes the main conclusions we draw from
296 these data. Nonetheless, future work could measure muscle activity or movement more directly,
297 such as with surface EMG of the masseter or articulatory tracking of the tongue, to resolve this
298 question more directly.

299 Overall, our study suggests that the direct updating of internal models through sensory prediction
300 errors is sufficient to drive speech adaptation. In our view, speech planning may provide the critical
301 sensory predictions that, when violated, lead to adaptation of internal models governing speech
302 control. These results do not support models of speech adaptation that rely solely on the
303 incorporation of feedback-based motor commands into future feedforward plans, suggesting that
304 such models could be revised to incorporate direct updating of internal models through sensory
305 error processing. Moreover, by extending our previous findings in upper-limb adaptation (Kim et
306 al., 2022) to a novel motor domain (speech) and a different sensory system (audition), we show

307 that plan-based predictions may form the basis for sensorimotor adaptation across a wide range
308 of human motor behaviors, pointing to a shared, domain-general neurocomputational mechanism.

309 Methods

310 Participants and power

311 30 participants were tested in the current study (5 male/25 female, age range 18-45, mean age
312 23.4). The sample size was determined using a bootstrapping procedure with effect sizes
313 observed in our previous work (Hantzsch et al., 2022), with the target of 90% power to detect a
314 similar sized effect at $\alpha = 0.05$. All participants were native speakers of American English, without
315 any reported history of neurological, speech, or hearing disorders. All participants passed an
316 automated Hughson-Westlake hearing screening (thresholds 25 dB HL or lower in both ears at
317 250, 500, 1000, 2000, and 4000 Hz). Participants gave written informed consent prior to
318 participation in the study and were compensated either monetarily or with course credit. All
319 procedures were approved by the Institutional Review Board of the University of Wisconsin-
320 Madison (protocol 2017-1128).

321 Task setup

322 Participants were seated in a sound-insulated booth in front of a computer monitor. On each trial,
323 a target word appeared on the screen in white text (Figure 1B). Each trial lasted 1.7 s from
324 stimulus onset. Trials were separated by 1.25 s plus a random delay of 0-0.5 s. Participants were
325 instructed to read the words as quickly as possible as they appeared on each trial. Two trial types
326 were used – "movement" trials and "no-movement" trials, following our previous study in reach
327 adaptation (Kim et al., 2022). On the majority of trials (movement trials), the word stayed white
328 for the duration of the trial. On a subset of trials (no-movement trials), the target word turned red
329 200 ms after it appeared and stayed red for the remainder of the trial. Participants were instructed
330 to not produce overt speech if and when the target word turned red. 200 ms was chosen after
331 pilot testing suggested that this delay results in the inhibition of speech on most, but not all, trials
332 in the majority of participants. This delay was therefore long enough to elicit movement planning
333 but short enough to enable mostly successful inhibition, thereby allowing us to test whether
334 sensory feedback given in the absence of overt movement could drive speech adaptation.

335 On movement trials, participants' speech was recorded (AKG C520), digitized with a USB sound
336 card (Focusrite Scarlett 2i2), processed through the Audapter software package (Cai et al., 2008;
337 Tourville et al., 2013), and played back to the participants over closed-back, over-the-ear
338 headphones (Beyerdynamic DT 770). Speech was played back at a volume of approximately 83
339 dB SPL and mixed with speech-shaped noise at approximately 60 dB SPL. The final level of the
340 playback speech signal varied with the amplitude of participants' produced speech. The noise,
341 combined with the closed-back headphones, served to minimize potential perception of the
342 participants' own unaltered speech, which may have otherwise been perceptible through air or
343 bone conduction. The latency of audio playback on our system is ~18 ms, as measured using the
344 protocol suggested by (Kim et al., 2020).

345 Trials were organized into “triplets.” The first and last trials of each triplet were always movement
346 trials with veridical auditory feedback (Figure 1C). On the middle trial of each triplet, the auditory
347 feedback participants received was always perturbed, such that the first vowel formant (F1, Figure
348 1A) was either raised or lowered by 125 mels (a perceptually-calibrated measurement of
349 frequency) throughout the trial using Audapter. Middle trials were either movement trials ($\frac{1}{3}$ of
350 triplets) or no-movement trials with a stop signal ($\frac{2}{3}$ of triplets). On perturbed movement trials, the
351 auditory feedback was perturbed in real time by identifying the vowel formants using linear
352 predictive coding (LPC) and filtering the speech signal to introduce a shift to those formants. On
353 perturbed no-movement trials, Audapter was used offline to apply the same shift to a recording of
354 the participants’ production from the previous trial (i.e. the first trial of the triplet, which was always
355 an unperturbed “movement” trial). On these trials, Audapter was initiated at the same time as on
356 movement trials, but played back this perturbed speech signal rather than playing back the
357 speech recorded in real time. Thus, the latency of the auditory feedback on each no-movement
358 trial was the same as on the immediately preceding trial.

359 Triplets were organized into blocks (Figure 1D). Each block contained 2 movement triplets and 4
360 no-movement triplets. Upward and downward frequency perturbations were balanced within each
361 block. The order of triplets within each block was randomized. Each block additionally contained
362 3 distractor trials, randomly inserted between triplets, in order to disrupt the rhythm of the
363 perturbations across trials. All distractor trials were movement trials with veridical auditory
364 feedback. Each block used a single stimulus word (“bed”, “dead”, or “Ted”, all sharing the same
365 target vowel / ϵ /); stimuli were pseudo-randomized across blocks such that there were an equal
366 number of blocks for all stimuli, and no consecutive blocks used the same stimulus. The
367 experiment consisted of 36 total blocks, yielding 216 total triplets. This resulted in 36 perturbations
368 in movement blocks and 72 perturbations in no-movement blocks for each perturbation direction.
369 A short self-timed break was allowed between each block.

370 In order to encourage participants to produce their speech movements quickly, they were given
371 points based on their response latency on movement trials. 10 points were given for responses
372 with latencies up to 500 ms, falling off by 1 point every 40 ms thereafter. To encourage participants
373 to inhibit overt speech production on no-movement trials, -25 points were given when a spoken
374 response was detected. On all trials, the onset of speech (and response latency) was determined
375 as the point where the amplitude of the microphone signal crossed above a predetermined, low
376 amplitude threshold.

377 Data quantification

378 Formant data were tracked using `wave_viewer` (Niziolek and Houde, 2015), which provides a
379 MATLAB GUI interface for formant tracking using Praat (Boersma and Weenink, 2019). LPC order
380 and pre-emphasis values were set individually for each participant. Vowels were initially
381 automatically identified by locating the samples which were above a participant-specific amplitude
382 level. Subsequently, all trials were hand-checked for errors. Errors in formant tracking were
383 corrected by adjusting the pre-emphasis value or LPC order. Errors in the location of vowel onset
384 and offset were corrected by hand-marking these times using landmarks in the audio waveform

385 and spectrogram. For each trial, F1 was averaged from 50-125 ms after vowel onset (Figure 1E)
386 to avoid 1) coarticulatory influences on vowel formants from the initial consonant and 2) potential
387 changes in vowel formants due to online feedback corrections, which begin roughly 150 ms after
388 vowel onset (Cai et al., 2012; Parrell et al., 2017). Recent work has shown that this window is the
389 most likely to accurately capture learning in this single-exposure paradigm (Hantzsch et al., 2022).

390 Adaptation was quantified in three ways. Single-trial adaptation was measured as the change in
391 F1 (in mels) from the first to the third trial of a triplet (i.e., after exposure to the auditory
392 perturbation). We additionally calculated single-trial aftereffects – the retention of learning – by
393 measuring the change in F1 (in mels) from the first trial in a triplet to the trial immediately following
394 the third trial of that triplet. Finally, we computed an “adaptation index,” which was quantified by
395 simply combining the above two metrics for each triplet. In all cases, the critical trials were
396 unperturbed movement trials, either the first trial of the following triplet or a distractor trial.
397 Because of the rather high variance in F1 production for vowels (standard deviation of ~15-20 Hz,
398 Whalen and Chen, 2019) relative to the maximum expected effect size given previous work (~2
399 Hz, Hantzsch et al., 2022), the adaptation index was our primary dependent variable of interest,
400 designed to maximize our statistical power and potential to detect a small effect relative to the
401 expected variance in production.

402 A small number of trials were excluded due to errors in production (i.e., the participant said the
403 wrong word), disfluencies, or unresolvable errors in formant tracking (0-3.6% across participants,
404 median .03%). Additionally, no-movement triplets where participants produced any detectable
405 vocalization on the middle perturbed trial were excluded in order to isolate learning effects when
406 no speech movement was produced (4-89 triplets excluded across participants, median 32). In
407 total, these exclusions resulted in 64-72 independent measurements of adaptation (median 71)
408 and 52-140 independent measurements of retention for each participant (median 111).

409 Statistical analysis

410 Following our previous study in reaching (Kim et al., 2022), linear mixed effects models were run
411 using the lme4 package in R (Bates et al., 2015). For single-trial adaptation, retention, and the
412 adaptation index models, model predictors included perturbation direction (upward or downward),
413 triplet type (movement or no-movement), and the interaction between these two factors.
414 Participant was included as a random effect (intercept). Statistical significance was assessed
415 using the lmerTest package (Kuznetsova et al., 2017). Post-hoc comparisons were conducted
416 with estimated marginal means using the Satterthwaite method for approximating the degrees of
417 freedom (Lenth et al., 2023). Effect sizes are reported as partial R^2 values from the linear mixed
418 effects models and Cohen's d for pairwise comparisons. Summary statistics report means and
419 standard errors.

420 Acknowledgments

421 Work supported by grants R01 DC019134 (C.A.N. & B.P.), R01 DC017091 (B.P.), and R01
422 NS132926 (S.D.M.) from the National Institutes of Health and BCS 2120506 (C.A.N. & B.P.) from
423 the National Science Foundation.

424 Competing Interests

425 The authors declare no competing interests.

426 References

- 427 Albert ST, Shadmehr R (2016) The Neural Feedback Response to Error As a Teaching Signal
428 for the Motor Learning System. *J Neurosci* 36:4832–4845.
- 429 Avraham G, Taylor JA, Breska A, Ivry RB, McDougale SD (2022) Contextual effects in
430 sensorimotor adaptation adhere to associative learning rules Lee T, Frank MJ, eds. *eLife*
431 11:e75801.
- 432 Babel M (2012) Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *J*
433 *Phon* 40:177–189.
- 434 Babel M, Bulatov D (2012) The Role of Fundamental Frequency in Phonetic Accommodation.
435 *Lang Speech* 55:231–248.
- 436 Bates D, Mächler M, Bolker B, Walker S (2015) Fitting Linear Mixed-Effects Models Using lme4.
437 *J Stat Softw* 67 Available at: <http://www.jstatsoft.org/v67/i01/> [Accessed May 15, 2020].
- 438 Beach SD, Tang D, Kiran S, Niziolek CA (2024) Pars Opercularis Underlies Efferent Predictions
439 and Successful Auditory Feedback Processing in Speech: Evidence From Left-
440 Hemisphere Stroke. *Neurobiol Lang* 5:454–483.
- 441 Boersma P, Weenink D (2019) Praat: doing phonetics by computer. Available at:
442 <http://www.praat.org/>.
- 443 Brudner SN, Kethidi N, Graeupner D, Ivry RB, Taylor JA (2016) Delayed feedback during
444 sensorimotor learning selectively disrupts adaptation but not strategy use. *J*
445 *Neurophysiol* 115:1499–1511.
- 446 Cai S, Beal DS, Ghosh SS, Tiede MK, Guenther FH, Perkell JS (2012) Weak responses to
447 auditory feedback perturbation during articulation in persons who stutter: evidence for
448 abnormal auditory-motor transformation. *PLoS One* 7:e41830.
- 449 Cai S, Boucek M, Ghosh S, Guenther FH, Perkell J (2008) A System for Online Dynamic
450 Perturbation of Formant Trajectories and Results from Perturbations of the Mandarin
451 Triphthong /iau/. In: *Proceedings of the 8th International Seminar on Speech Production*,
452 pp 65–68. Strasbourg, France.
- 453 Caudrelier T, Schwartz J-L, Perrier P, Gerber S, Rochet-Capellan A (2018) Transfer of
454 Learning: What Does It Tell Us About Speech Production Units? *J Speech Lang Hear*
455 *Res*:1.
- 456 Churchland MM, Shenoy KV (2024) Preparatory activity and the expansive null-space. *Nat Rev*
457 *Neurosci* 25:213–236.
- 458 Curio G, Neuloh G, Numminen J, Jousmäki V, Hari R (2000) Speaking modifies voice-evoked
459 activity in the human auditory cortex. *Hum Brain Mapp* 9:183–191.
- 460 Daliri A, Max L (2016) Modulation of Auditory Responses to Speech vs. Nonspeech Stimuli
461 during Speech Movement Planning. *Front Hum Neurosci* 10:234.
- 462 Day KA, Roemmich RT, Taylor JA, Bastian AJ (2016) Visuomotor Learning Generalizes Around

- 463 the Intended Movement. *eNeuro* 3 Available at:
464 <https://www.eneuro.org/content/3/2/ENEURO.0005-16.2016> [Accessed April 5, 2024].
- 465 Donchin O, Francis JT, Shadmehr R (2003) Quantifying Generalization from Trial-by-Trial
466 Behavior of Adaptive Systems that Learn with Basis Functions: Theory and Experiments
467 in Human Motor Control. *J Neurosci* 23:9032–9045.
- 468 Ebrahimi S, Ostry DJ (2024) The human somatosensory cortex contributes to the encoding of
469 newly learned movements. *Proc Natl Acad Sci* 121:e2316294121.
- 470 Eliades SJ, Wang X (2003) Sensory-Motor Interaction in the Primate Auditory Cortex
471 During Self-Initiated Vocalizations. *J Neurophysiol* 89:2194–2207.
- 472 Flinker A, Chang EF, Kirsch HE, Barbaro NM, Crone NE, Knight RT (2010) Single-Trial Speech
473 Suppression of Auditory Cortex Activity in Humans. *J Neurosci* 30:16643–16650.
- 474 Gandolfo F, Mussa-Ivaldi FA, Bizzi E (1996) Motor learning by field approximation. *Proc Natl*
475 *Acad Sci* 93:3843–3846.
- 476 Goldinger SD (1998) Echoes of echoes? An episodic theory of lexical access. *Psychol Rev*
477 105:251–279.
- 478 Guenther FH (2016) Neural control of speech. Cambridge, MA: The MIT Press.
- 479 Hadjosif AM, Krakauer JW, Haith AM (2021) Did We Get Sensorimotor Adaptation Wrong?
480 Implicit Adaptation as Direct Policy Updating Rather than Forward-Model-Based
481 Learning. *J Neurosci* 41:2747–2761.
- 482 Hantzsch L, Parrell B, Niziolek CA (2022) A single exposure to altered auditory feedback
483 causes observable sensorimotor adaptation in speech. *eLife* 11:e73694.
- 484 Heald JB, Ingram JN, Flanagan JR, Wolpert DM (2018) Multiple motor memories are learned to
485 control different points on a tool. *Nat Hum Behav* 2:300–311.
- 486 Heald JB, Lengyel M, Wolpert DM (2021) Contextual inference underlies the learning of
487 sensorimotor repertoires. *Nature*:1–5.
- 488 Houde JF, Nagarajan SS (2011) Speech production as state feedback control. *Front Hum*
489 *Neurosci* 5:82.
- 490 Houde JF, Nagarajan SS, Sekihara K, Merzenich MM (2002) Modulation of the Auditory Cortex
491 during Speech: An MEG Study. *J Cogn Neurosci* 14:1125–1138.
- 492 Howard IS, Wolpert DM, Franklin DW (2013) The effect of contextual cues on the encoding of
493 motor memories. *J Neurophysiol* 109:2632–2644.
- 494 Kawato M, Furukawa K, Suzuki R (1987) A hierarchical neural-network model for control and
495 learning of voluntary movement. *Biol Cybern* 57:169–185.
- 496 Kearney E, Nieto-Castañón A, Weerathunge HR, Falsini R, Daliri A, Abur D, Ballard KJ, Chang
497 S-E, Chao S-C, Heller Murray ES, Scott TL, Guenther FH (2020) A Simple 3-Parameter
498 Model for Examining Adaptation in Speech and Voice Production. *Front Psychol* 10
499 Available at:
500 <https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2019.02995/full>
501 [Accessed April 4, 2024].
- 502 Kim K, Wang H, Max L (2020) It's About Time: Minimizing Hardware and Software Latencies in
503 Speech Research with Real-Time Auditory Feedback. *J Speech Lang Hear Res*
504 63:2522–2534.
- 505 Kim KS, Gaines JL, Parrell B, Ramanarayanan V, Nagarajan SS, Houde JF (2023a)
506 Mechanisms of sensorimotor adaptation in a hierarchical state feedback control model of
507 speech. *PLOS Comput Biol* 19:e1011244.
- 508 Kim KS, Hinkley LB, Dale CL, Nagarajan SS, Houde JF (2023b) Neurophysiological evidence of
509 sensory prediction errors driving speech sensorimotor adaptation. :2023.10.22.563504
510 Available at: <https://www.biorxiv.org/content/10.1101/2023.10.22.563504v1> [Accessed
511 October 27, 2023].
- 512 Kim OA, Forreance AD, McDougle SD (2022) Motor learning without movement. *Proc Natl Acad*

- 513 Sci 119:e2204379119.
- 514 Kitazawa S, Kohno T, Uka T (1995) Effects of delayed visual information on the rate and
515 amount of prism adaptation in the human. *J Neurosci* 15:7644–7652.
- 516 Krakauer JW, Hadjiosif AM, Xu J, Wong AL, Haith AM (2019) Motor Learning. *Compr Physiol*
517 9:613–663.
- 518 Krakauer JW, Pine ZM, Ghilardi M-F, Ghez C (2000) Learning of Visuomotor Transformations
519 for Vectorial Planning of Reaching Trajectories. *J Neurosci* 20:8916–8924.
- 520 Kuznetsova A, Brockhoff PB, Christensen RHB (2017) lmerTest Package: Tests in Linear Mixed
521 Effects Models. *J Stat Softw* 82 Available at: <http://www.jstatsoft.org/v82/i13/> [Accessed
522 May 15, 2020].
- 523 Lenth R, Singmann H, Love J, Buerkner P, Herve M (2023) emmeans: Estimated Marginal
524 Means, aka Least-Squares Means. Available at: [https://CRAN.R-](https://CRAN.R-project.org/package=emmeans)
525 [project.org/package=emmeans](https://CRAN.R-project.org/package=emmeans).
- 526 Mattar AAG, Gribble PL (2005) Motor Learning by Observing. *Neuron* 46:153–160.
- 527 Max L, Maffett DG (2015) Feedback delays eliminate auditory-motor learning in speech
528 production. *Neurosci Lett* 591:25–29.
- 529 McDougle SD, Bond KM, Taylor JA (2017) Implications of plan-based generalization in
530 sensorimotor adaptation. *J Neurophysiol:jn.00974.2016*.
- 531 Niziolek CA, Houde J (2015) Wave_Viewer: First Release. Available at:
532 <https://zenodo.org/record/13839> [Accessed January 30, 2020].
- 533 Niziolek CA, Nagarajan SS, Houde JF (2013) What does motor efference copy represent?
534 Evidence from speech production. *J Neurosci* 33:16110–16116.
- 535 Parrell B, Agnew Z, Nagarajan S, Houde JF, Ivry RB (2017) Impaired Feedforward Control and
536 Enhanced Feedback Control of Speech in Patients with Cerebellar Degeneration. *J*
537 *Neurosci* 37:9249–9258.
- 538 Parrell B, Ramanarayanan V, Nagarajan S, Houde J (2019) The FACTS model of speech motor
539 control: Fusing state estimation and task-based control. *PLOS Comput Biol*
540 15:e1007321.
- 541 Pawlowsky C, Thénault F, Bernier P-M (2023) Implicit Sensorimotor Adaptation Proceeds in
542 Absence of Movement Execution. *eNeuro* 10 Available at:
543 <https://www.eneuro.org/content/10/8/ENEURO.0508-22.2023> [Accessed April 8, 2024].
- 544 Rochet-Capellan A, Richer L, Ostry DJ (2012) Nonhomogeneous transfer reveals specificity in
545 speech motor learning. *J Neurophysiol* 107:1711–1717.
- 546 Schween R, Hegele M (2017) Feedback delay attenuates implicit but facilitates explicit
547 adjustments to a visuomotor rotation. *Neurobiol Learn Mem* 140:124–133.
- 548 Scott SH (2004) Optimal feedback control and the neural basis of volitional motor control. *Nat*
549 *Rev Neurosci* 5:532–546.
- 550 Shadmehr R, Krakauer JW (2008) A computational neuroanatomy for motor control. *Exp Brain*
551 *Res* 185:359–381.
- 552 Shadmehr R, Smith MA, Krakauer JW (2010) Error correction, sensory prediction, and
553 adaptation in motor control. *Annu Rev Neurosci* 33:89–108.
- 554 Sheahan HR, Franklin DW, Wolpert DM (2016) Motor Planning, Not Execution, Separates Motor
555 Memories. *Neuron* 92:773–779.
- 556 Sheahan HR, Ingram JN, Żalalyté GM, Wolpert DM (2018) Imagery of movements immediately
557 following performance allows learning of motor skills that interfere. *Sci Rep* 8:14330.
- 558 Shiller DM, Mitsuya T, Max L (2020) Exposure to Auditory Feedback Delay while Speaking
559 Induces Perceptual Habituation but does not Mitigate the Disruptive Effect of Delay on
560 Speech Auditory-motor Learning. *Neuroscience* 446:213–224.
- 561 Shockley K, Sabadini L, Fowler CA (2004) Imitation in shadowing words. *Percept Psychophys*
562 66:422–429.

- 563 Smith MA, Ghazizadeh A, Shadmehr R (2006) Interacting Adaptive Processes with Different
564 Timescales Underlie Short-Term Motor Learning Ashe J, ed. PLoS Biol 4:e179.
- 565 Tang D, Parrell B, Beach SD, Niziolek CA (2023) The brain's sensitivity to sensory error can be
566 modulated by altering perceived variability. :2023.06.26.546615 Available at:
567 <https://www.biorxiv.org/content/10.1101/2023.06.26.546615v1> [Accessed February 19,
568 2024].
- 569 Tilsen S, Goldstein L (2012) Articulatory gestures are individually selected in production. J Phon
570 40:764–779.
- 571 Tourville JA, Cai S, Guenther F (2013) Exploring auditory-motor interactions in normal and
572 disordered speech. In: Proceedings of Meetings on Acoustics, pp 060180 Available at:
573 <http://asa.scitation.org/doi/abs/10.1121/1.4800684> [Accessed April 9, 2020].
- 574 Tourville JA, Guenther FH (2011) The DIVA model: A neural theory of speech acquisition and
575 production. Lang Cogn Process 26:952–981.
- 576 Tsay JS, Kim H, Haith AM, Ivry RB (2022) Understanding implicit sensorimotor adaptation as a
577 process of proprioceptive re-alignment Pruszynski JA, Behrens TE, eds. eLife
578 11:e76639.
- 579 Ventura MI, Nagarajan SS, Houde JF (2009) Speech target modulates speaking induced
580 suppression in auditory cortex. BMC Neurosci 10:58.
- 581 Vyas S, O'Shea DJ, Ryu SI, Shenoy KV (2020) Causal Role of Motor Preparation during Error-
582 Driven Learning. Neuron 106:329-339.e4.
- 583 Whalen DH, Chen W-R (2019) Variability and Central Tendencies in Speech Production. Front
584 Commun 4:49.
- 585 Wolpert DM, Diedrichsen J, Flanagan JR (2011) Principles of sensorimotor learning. Nat Rev
586 Neurosci 12:739–751.
- 587 Wolpert DM, Flanagan JR (2016) Computations underlying sensorimotor learning. Curr Opin
588 Neurobiol 37:7–11.
- 589 Wolpert DM, Kawato M (1998) Multiple paired forward and inverse models for motor control.
590 Neural Netw 11:1317–1329.
- 591 Wolpert DM, Miall RC, Kawato M (1998) Internal models in the cerebellum. Trends Cogn Sci
592 2:338–347.
- 593 Zhou W, Fitzgerald J, Colucci-Chang K, Murthy KG, Joiner WM (2017) The temporal stability of
594 visuomotor adaptation generalization. J Neurophysiol 118:2435–2447.