Neurophysiological evidence of sensory prediction errors driving speech sensorimotor adaptation

Kwang S. Kim[1], Leighton B. Hinkley[2], Corby L. Dale[2], Srikantan S. Nagarajan[2], and John F. Houde[3]

1. Department of Speech, Language, and Hearing Sciences, Purdue University, West Lafayette, IN
2. Department of Radiology and Biomedical Imaging, University of California San Francisco, San Francisco, CA
3. Department of Otolaryngology—Head and Neck Surgery, University of California San Francisco, San Francisco, CA

Correspondence: Kwang S. Kim, 715 Clinic Drive (Lyles-Porter Hall), West Lafayette, IN 47907, USA. Email: kwangkim@purdue.edu

Acknowledgement: This study was funded by grants from the National Institute of Health (F32DC019538 to K.S.K., R01DC017696, R01DC017091, R01DC013979, R01NS100440, and P50DC019900 to J.F.H., S.S.N.). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. We would also like to thank Ashley Tay, Joshua Chon, Mahmoud Jiha, Derek Kinsella, Gavin Belok, and Lingwei Ouyang for their help with data collection and extraction.

Abstract

The human sensorimotor system has a remarkable ability to quickly and efficiently learn movements from sensory experience. A prominent example is sensorimotor adaptation, learning that characterizes the sensorimotor system's response to persistent sensory errors by adjusting future movements to compensate for those errors. Despite being essential for maintaining and fine-tuning motor control, mechanisms underlying sensorimotor adaptation remain unclear. A component of sensorimotor adaptation is implicit (i.e., the learner is unaware of the learning process) which has been suggested to result from sensory prediction errors–the discrepancies between predicted sensory consequences of motor commands and actual sensory feedback. However, to date no direct neurophysiological evidence that sensory prediction errors drive adaptation has been demonstrated. Here, we examined prediction errors via magnetoencephalography (MEG) imaging of the auditory cortex during sensorimotor adaptation of speech to altered auditory feedback, an entirely implicit adaptation task. Specifically, we measured how speaking-induced suppression (SIS)--a neural representation of auditory prediction errors--changed over the trials of the adaptation experiment. SIS refers to the suppression of auditory cortical response to speech onset (in particular, the M100 response) to self-produced speech when compared to the response to passive listening to identical playback of that speech. SIS was reduced (reflecting larger prediction errors) during the early learning phase compared to the initial unaltered feedback phase. Furthermore, reduction in SIS positively correlated with behavioral adaptation extents, suggesting that larger prediction errors were associated with more learning. In contrast, such a reduction in SIS was not found in a control experiment in which participants heard unaltered feedback and thus did not adapt. In addition, in some participants who reached a plateau in the late learning phase, SIS increased (reflecting smaller prediction errors), demonstrating that prediction errors were minimal when there was no further adaptation. Together, these findings provide the first neurophysiological evidence for the hypothesis that prediction errors drive human sensorimotor adaptation.

Introduction

The sensorimotor system shows a remarkable ability to quickly and efficiently learn movements based on sensory feedback. Soon after perceiving sensory errors that arise from movements, the system updates future movements to compensate for the errors, a phenomenon called sensorimotor adaptation. What drives such an elegant learning process? Previous studies suggested that adaptation can be driven by both task errors (i.e., discrepancy between the action and the goal) and sesnory prediction errors (i.e., mismatches between the actual sensory consequences of a movement and those predicted from the motor commands driving that movement).

In the speech domain, however, multiple lines of evidence suggest that speech sensorimotor adaptation to altered auditory feedback is implicit (i.e., participants are unaware of the learning), and hypothesized to be driven mainly by sensory prediction errors (Mazzoni & Krakauer, 2006). For example, participants showed no difference in the amount of learning in response to formant-perturbed auditory feedback when instructed to compensate, to ignore the feedback, or to avoid compensating (Keough et al., 2013; Munhall et al., 2009). Although behavioral studies have suggested that this unconscious minimizing of auditory prediction errors is the signal that drives speech sensorimotor adaptation, direct neurophysiological evidence of this process has not been demonstrated.

A neural representation of auditory prediction errors is speaking-induced suppression (SIS) of the auditory cortex. Studies have reported that the auditory responses to self-produced speech are smaller (i.e., suppressed) than the responses to playback of the same speech sound, consistent with the idea that auditory responses arise from auditory prediction errors, which are small in the self-produced case (i.e., auditory feedback is predictable) and large in the passively heard case (i.e., auditory feedback is unpredictable). Thus, SIS demonstrates that, during speaking, the auditory system predicts and anticipates the arrival of auditory feedback of speech onset, resulting in a suppressed feedback comparison response, as compared to auditory responses during passive listening to playback when speech onset cannot be predicted/anticipated. Consistent with the idea, SIS was reduced when participants spoke with pitch-perturbed auditory feedback (e.g., Behroozmand & Larson, 2011; Chang et al., 2013) or voice-manipulated auditory feedback ("alien voice", e.g., Heinks-Maldonado et al., 2005, 2006; Houde et al., 2002). Importantly, this reduction in the suppression of auditory areas in response to perturbed auditory feedback are not unique to human speech, as they have also been observed in marmoset monkey vocal production (e.g., Eliades & Tsunada, 2018).

Previously, reduction in a similar suppression effect (i.e., suppressed neural response in active movements compared to passive movements) has been found in Rhesus monkey cerebellum during sensorimotor adaptation (Brooks et al., 2015), but no such evidence has been documented in humans to date. One previous study that examined SIS during adaptation to first formant frequency shifts via electroencephalography (EEG) reported that SIS amplitude in the learning phase (i.e., during perturbed first formant) was not reduced compared to the pre-adaptation baseline (Sato & Shiller, 2018). However, the negative finding could result from masking of SIS changes across all 80 feedback perturbation trials, as opposed to changes that may have occurred in early trials (e.g., initial 20 to 40 feedback perturbation trials) when most adaptation occurs (e.g., Kim & Max, 2021). Here, we used magnetoencephalography (MEG) imaging during repeated speech adaptation sessions to test the hypotheses that (1) SIS reduces during early phases of speech sensorimotor adaptation, and (2) the early SIS reduction may be distinct from SIS changes found in later phases of adaptation.

Results

Participants lay supine on the scanner bed of a whole-head, 275-channel biomagnetometer system (MEG; Omega 2000, CTF, Coquitlam, BC, Canada) for a total of four sessions (first and second speaking sessions, first and second listening sessions). During the first two sessions, participants were asked to read “Ed,” “end,” or “ebb” (60 trial blocks for 3 different words = 180 total trials) that appeared on the screen. During these speaking sessions, participants heard their speech with the first formant frequency (Formant 1 or F1) shifted upward for some trials, which made their speech to sound like “Add,” “And,” and “Ab,” respectively. Specifically, after the first 20 trial blocks (i.e., baseline) which had no perturbation, the 150 Hz up-shift perturbation was present from trial block 21 to 50. We categorized the first 15 trial blocks of the perturbed trials (21 – 45) as the early learning phase and the second 15 trial blocks (36-50) as the late learning phase.

After the first session, participants were given a few minute-long break that included conversations with the experimenter, which allowed additional exposure to their unaltered auditory feedback (Figure 1). We then asked participants to repeat another speaking session. The rationale for this repeated session was that most adaptation occurs quickly, often in the first 10-30 trials of the perturbation phase, but such a low number of trials does not provide enough power for the evoked potential analyses. Thus, to ensure an adequate number of trials for the early and late learning phases, an additional session was recorded. After completing two speaking sessions, participants were asked to listen to their recorded speech in the first two speaking sessions across the subsequent two sessions (i.e., listening sessions). During the listening sessions, participants saw the same stimuli (i.e., words) that they saw in the speaking sessions (see Methods for more details).

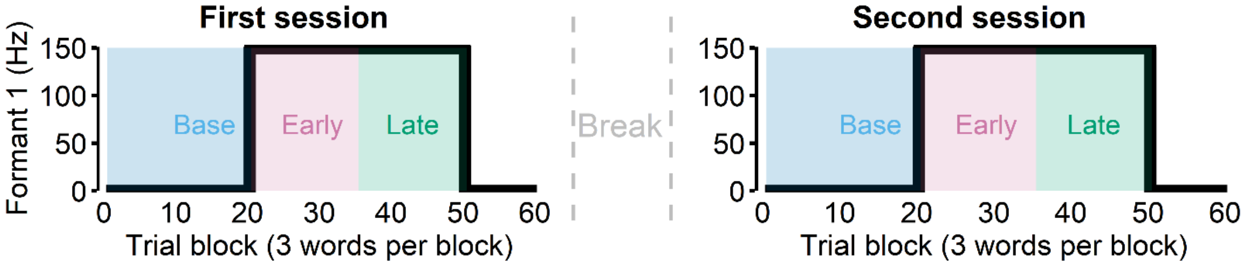

Figure 1. Participants were asked to read words during the first two sessions ("speak"). In these sessions, 150 Hz up-shift perturbation was present from the trial block 21 to 50. We categorized the first 15 trial blocks of the perturbed trials (21 – 45) as the early learning phase and the second 15 trial blocks (36-50) as the late learning phase. After the first session, we asked participants to repeat another speaking session after a few minute-long break.

We averaged the acoustic and MEG data across the repeated sessions. As shown in Figure 2, source localization of trial-averaged data for each condition (speak, listen) and phase (baseline, early learning, and late learning) was conducted to determine peak activity (M100) location within the auditory cortex. We then computed the M100 amplitude differences between the listen and speak sessions to determine SIS for each condition and phase (see Methods for more details).

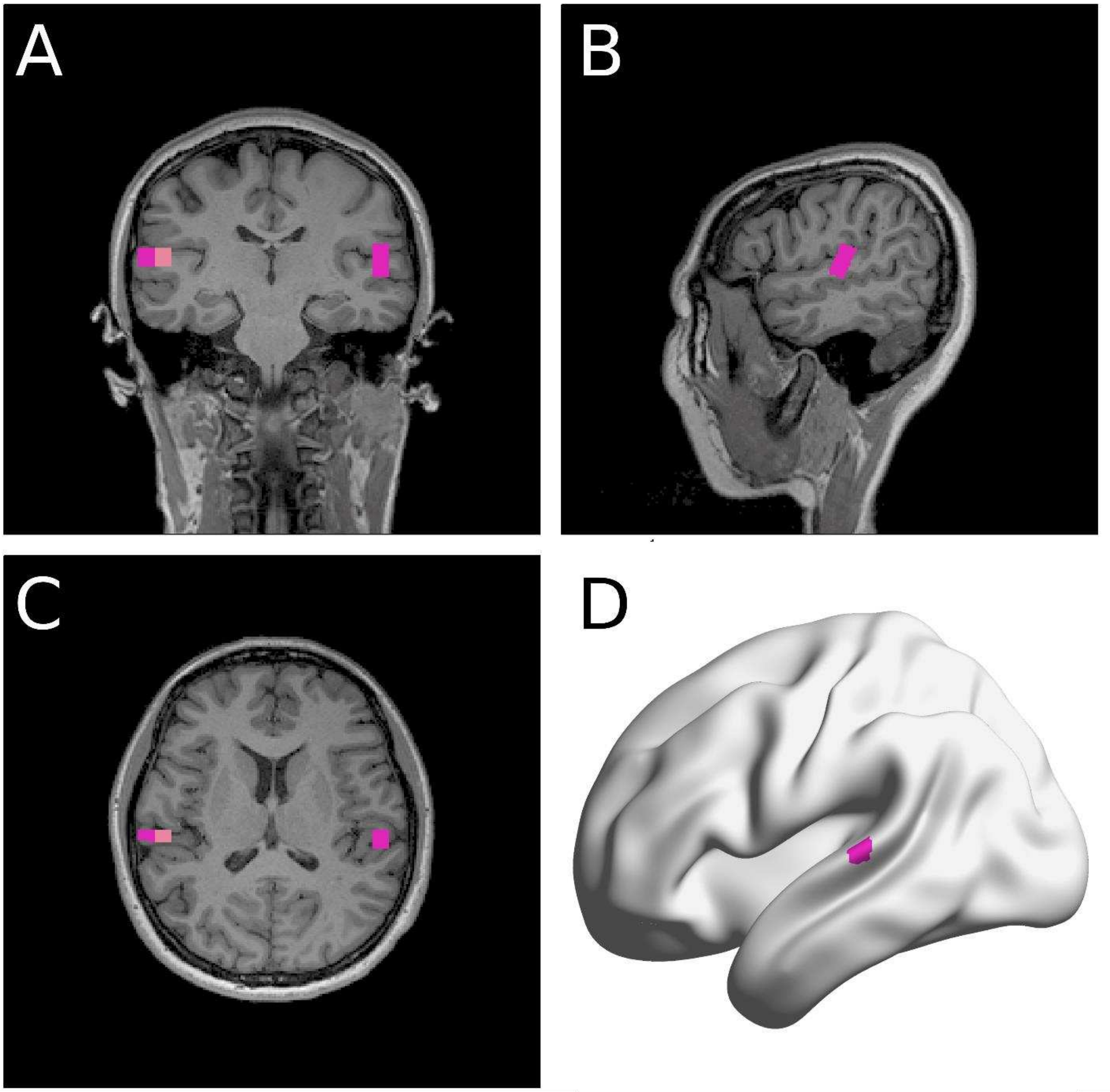

Figure 2. A representative participant's source localization. NUTMEG (citation) identified a few MNI coordinates that showed clear M100 response shown in the coronal (**A**), sagittal (**B**), and transverse (**C**) planes. The MNI coordinate of the voxel with the most power in the auditory areas in each hemisphere was selected for analyses. **D**: The same participant's for left auditory area coordinate selected shown in a surface-based rendering (BrainNet Viewer, Xia et al., 2013).

SIS was reduced during early learning and the reduction was positively correlated with adaptation

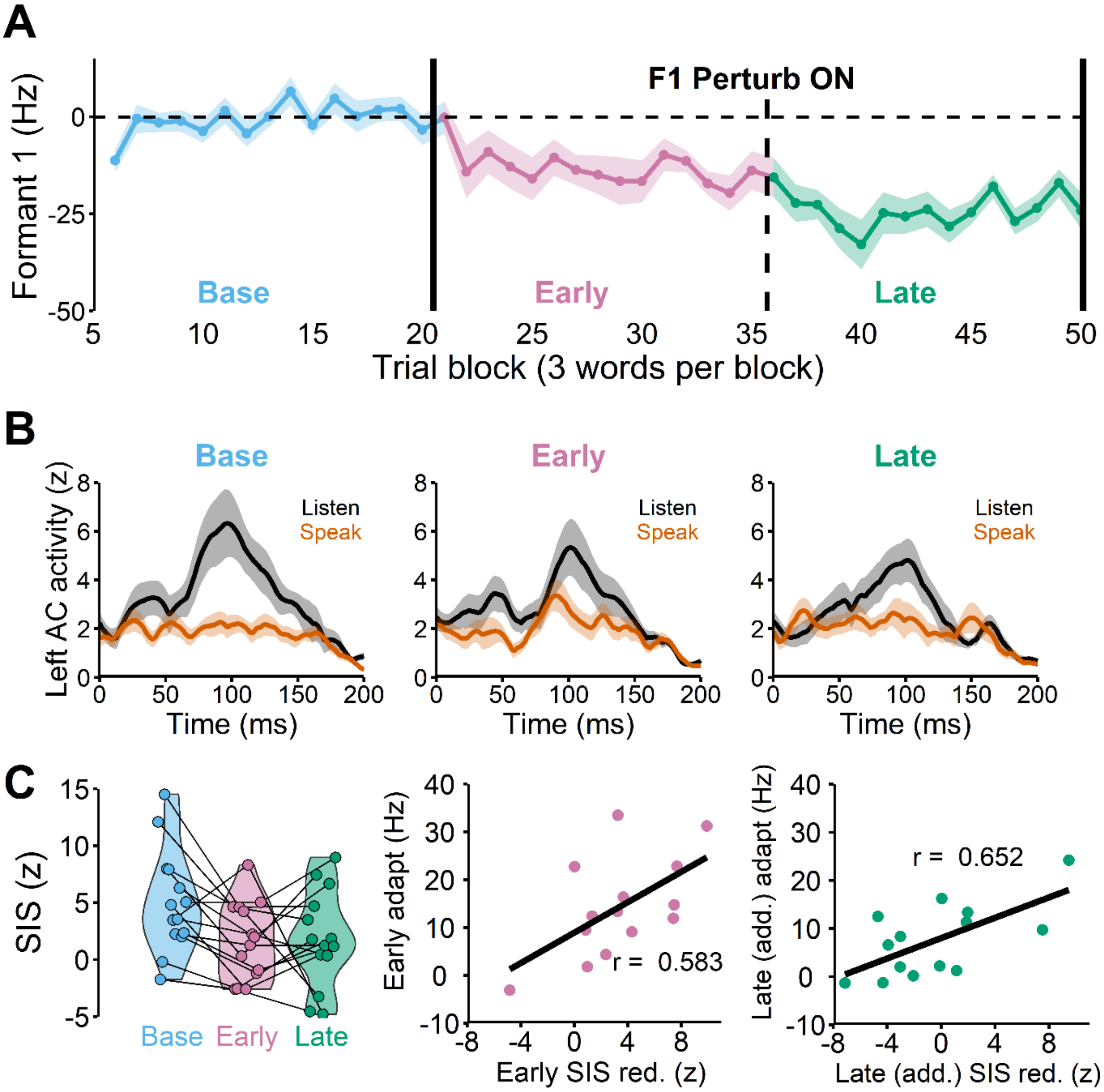

Figure 3 **A**: The group average speech auditory-motor adaptation in which participants lowered their first formant frequency (F1) in response to the 150 Hz upshift F1 perturbation. **B**: The left auditory cortex responses (M100) in listen and speak conditions demonstrate that the amount of speaking-induced suppression (i.e., listen (black) – speak (orange)) is reduced during early learning (Early) compared to the baseline (Base). **C**: SIS was significantly reduced in the early and late learning phases compared to the baseline (left, r(12) = 0.583, p = 0.029). The amount of SIS reduction in the early learning phase was significantly correlated with the amount of early adaptation (middle). The amount of additional SIS reduction in the late learning phase also significantly correlated with the additional amount of adaptation in the phase (right, r(12) = 0.652, p = 0.011).

Nearly all participants adapted in both speaking sessions (Fig. 3A), except for three participants who adapted in only one of the two sessions. Given that there was no evidence of savings (i.e., changes in the baseline or learning behavior from repeating the task, see Supplemental Information 1), these participants were included in the analyses. The SIS analyses revealed that there was no right hemisphere SIS (see Supplemental Information 2), which is known to be variable across tasks and individuals (see Discussion for more details). On the other hand, most participants showed a clear suppression of left auditory activity in the speaking condition (compared to the listening condition) during the baseline phase (Fig. 3B, left). Hence, SIS refers to suppression of *left* auditory activity hereafter unless specified otherwise.

We also found that the SIS response changed in the early and late learning phases (Fig. 3B, middle and right), F(2, 28) = 5.131, p = 0.013. The post-hoc pairwise comparison test indicated that SIS response was significantly reduced in the early learning phase compared to the baseline (Fig. 3C, left), t(30.1) =2.749, p_adjusted = 0.026, demonstrating that there were large auditory prediction errors during the early learning phase. Additionally, we found that the amount of SIS reduction in the early learning phase was positively correlated with the amount of learning (in the early learning phase) across participants, r(12) = 0.583, p = 0.029 (Fig. 3C, middle).

Further SIS reduction was positively correlated with (additional) late learning

The SIS amplitude in the late learning phase was also significantly reduced compared to the baseline (Fig. 3C, left), t(30.1) = 2.591, p_adjust = 0.038. Importantly, we found that the SIS reduction from the baseline was not significantly correlated with the final amount of adaptation in the late learning phase, r(12) = 0.260, p = 0.370. This result was consistent with our hypothesis that most learning typically occurs in the early phase, and thus the late phase SIS reduction from baseline would not be able to capture most of the adaptation extent. Rather, late SIS reduction that accounts for early SIS changes (i.e., *additional* late SIS reduction from early SIS) is likely a predictor for late (additional) learning behaviors. Indeed, we found that additional SIS reduction in the late learning phase (i.e., late SIS relative to the early SIS) was significantly correlated with additional late adaptation, i.e., late adaptation relative to early adaptation r(12) = 0.689, p = 0.001.

It should be noted that one participant with the largest additional SIS reduction (and the largest additional adaptation) had a big impact on the correlation. The participant's SIS reduction was indeed an outlier (outside of the 1.5 inter-quartile range above 75% percentile). After excluding the participant from the analysis, although the correlation was still positive, r(11) = 0.430, it was no longer statistically significant, p = 0.142. Nonetheless, we included this participant's data in this analysis because the SIS reduction for the participant (~12 z) was not too far from the maximum SIS reduction value found in the early learning data (~10 z). In fact, the participant's data would not have been considered an outlier if it was found in the early learning phase. Additionally, when we imputed the participant's data by replacing it with the 95th percentile of the group, the correlation was still statistically significant, r(12) = 0.652, p = 0.011, (see Fig. 3C, right). Lastly, a visual inspection of the participant's behavioral data clearly indicated large additional learning that was not due to outliers in the adaptation data, consistent with the participant's large SIS reduction (see Supplemental Information 3).

Another interesting finding is that there were 8 participants whose SIS increased in the learning phase, which resulted in a near-complete SIS recovery (i.e., the late learning SIS response did not differ from the baseline SIS response, t(7) = 0.824, p = 0.437. Importantly, these participants also did not show a significant amount of additional learning in this phase (compared to the early adaptation), t(7) = 2.082, p = 0.076 even though adaptation was largely incomplete (i.e., 14.88% of the perturbation size). Taken together, the relationship between additional SIS reduction and adaptation in the late learning phase also followed the same trend found in the early learning phase. That is, individuals who showed more reduction in SIS, also tended to show more learning, suggesting that larger adaptation was associated with larger prediction errors. In contrast, less learning or no learning behavior (e.g., reaching a plateau) was associated with smaller prediction errors (i.e., increases in SIS).

SIS remained unchanged when there was no learning

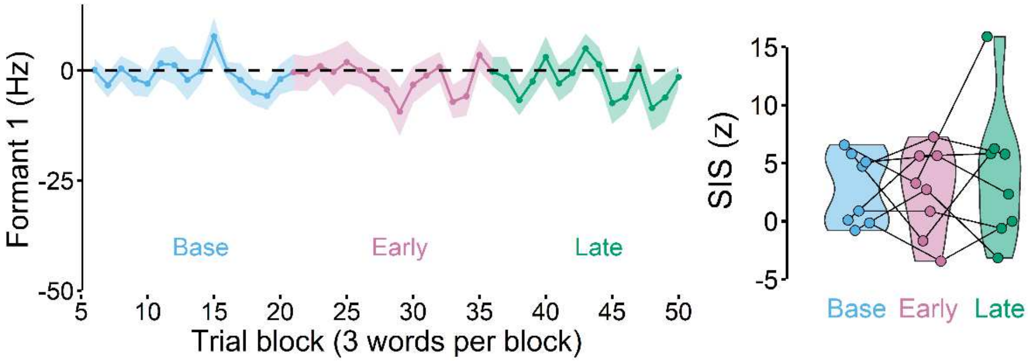

Figure 4. A control experiment in which no auditory perturbation was applied. As expected, participants did not show any changes in Formant 1, exhibiting, no learning (left). There was also no SIS change across the different phases (right).

To ensure that SIS reduction was related to learning behaviors, we designed a control experiment in which there was no auditory perturbation (and thus no learning was expected). Here, participants also completed two speaking and two listening sessions. Other than the absence of the perturbation, the experimental setup and the analyses methods were identical to the main experiment. We found that participants did not adapt (Fig 4A) and SIS reduction also did not occur (i.e., SIS amplitudes did not change across the phases), F(2, 16) = 0.484, p = 0.625. Therefore, SIS remained unchanged when there was no learning.

Discussion

We used magnetoencephalography (MEG) imaging to examine auditory prediction errors during speech auditory-motor adaptation. Specifically, we measured Speaking-Induced Suppression (SIS)—suppression of auditory responses to self-produced speech compared to the responses to passively heard speech—is thought to represent auditory prediction errors. To fully capture SIS changes in the early learning phase during which most of adaptation typically occurs, we analyzed the early learning and late learning phases separately.

Neurophysiological evidence that auditory prediction errors drive implicit adaptation

SIS was significantly reduced in the early learning phase during which adaptation occurred. In contrast, in a control experiment in which there was no perturbation (and thus no adaptation), such a SIS reduction was not found. In addition, the amount of SIS reduction was positively correlated with the amount of adaptation, delineating a direct link between prediction errors (i.e., more SIS reduction) and adaptation. Together, our findings demonstrate that auditory prediction errors drive speech auditory-motor adaptation. Our findings are consistent with previous reports of speech adaptation being entirely implicit (e.g., Kim & Max, 2021; Lametti et al., 2020), which is thought to be driven by prediction errors (Haith & Krakauer, 2013; Mazzoni & Krakauer, 2006). In addition, speech adaptation also seems to be sensitive to auditory feedback delays (i.e., 100 ms delay can eliminate adaptation), which highlights the importance of prediction errors that require temporally precise comparison of prediction and the actual feedback (Max & Maffett, 2015; Shiller et al., 2020). More recently, a computational model, Feedback-Aware Control of Tasks in Speech (FACTS, Parrell et al., 2019) also generated simulations of adaptation driven by auditory prediction errors (K. S. Kim et al., 2023).

To date, only one other study examined SIS during speech auditory-motor adaptation, but they reported no SIS changes during adaptation (Sato & Shiller, 2018). Although their finding may seem contradictory to the current study at first glance, it should be noted that in the previous study SIS amplitudes across the whole learning phase (80 trials) were averaged and analyzed together, which likely included SIS recovery response in the late phase as found in the current study's late learning phase. Hence, it is possible that SIS reduction was present in the early learning phase, but such an effect may have been weakened by the late perturbation data.

It should be noted that our findings do not necessarily reject the notion that task errors may also drive implicit speech adaptation. In upper limb visuomotor rotation, recent studies have demonstrated that task errors contribute to implicit adaptation (Albert et al., 2022; H. E. Kim et al., 2019; Leow et al., 2018, 2020; Miyamoto et al., 2020; Morehead & Xivry, 2021). Although it remains possible that other types of errors (in addition to prediction errors) may also influence speech adaptation, such evidence has not been documented (also see "What does SIS reflect?" below).

Broadly, our findings provide the first neurophysiological evidence that prediction errors drive implicit adaptation in humans. Previously a similar suppression effect has been documented in the cerebellum of rhesus monkey during head movement adaptation (Brooks et al., 2015). In the study, cerebellar neuron activities, which are typically suppressed during voluntary movements compared to passive movements much like SIS, did not differ between the two conditions (voluntary vs. passive) during adaptation. Remarkably, this reduced suppression also recovered (i.e., suppression increased) towards later learning trials, directly in line with our result. Here, we expanded the previous finding by demonstrating that the extent of such suppression reduction (or recovery) was closely associated with implicit adaptation across individuals.

Adaptation plateaus when prediction errors are minimal

Another interesting aspect of our finding reveals a potential mechanism that causes adaptation to halt. In the past, several explanations for why adaptation is incomplete have been put forth, especially for speech adaptation which often plateaus around 20-40% (see Kitchen et al., 2022 for detailed discussion). Some have demonstrated that speech adaptation accompanies changes in perceptual boundaries which may contribute to incomplete adaptation (Lametti et al., 2014; Shiller et al., 2009), but perceptual auditory targets do not seem to change throughout adaptation (K. S. Kim & Max, 2021) and preventing perceptual target shifts by playing back the participants' baseline productions did not increase adaptation. Others argued that a conflict between unperturbed somatosensory feedback and perturbed auditory feedback may lead to limited adaptation, but this account also lacks supporting evidence. In fact, preliminary data from our laboratory shows that even when somatosensory feedback becomes unreliable by oral application of lidocaine, adaptation behavior does not increase, suggesting that somatosensory feedback may not be a reason for incomplete adaptation.

An idea that has been proposed by an upper limb reaching adaptation study is that consistency of errors modulates error sensitivity, which results in limited adaptation (e.g., Albert et al., 2021). This idea has not been directly examined in the context of speech adaptation, but it is plausible that the overall size of prediction errors may be modulated by feedback (or perturbation) consistency. Some studies have found that individuals with high perceptual (auditory) acuity measured by psychometric functions had a larger extent of adaptation (e.g., Daliri & Dittman, 2019), which may suggest a potential link between error sensitivity and adaptation. Nonetheless, several studies also documented no such relationship (e.g., Abur et al., 2018; Alemi et al., 2021; Feng et al., 2011; Lester-Smith et al., 2020).

Recently, another potential explanation, which is that adaptation is halted by prediction errors which quickly decrease throughout adaptation because of both the motor output changes and sensory prediction updates, has been put forth by a computational model, FACTS (K. S. Kim et al., 2023). In the simulation, the adaptive motor output produced lower F1 in response to F1 upshift perturbation, resulting in perturbed sensory feedback to become more like the baseline sensory feedback (i.e., lower perturbed feedback in F1). Interestingly, the simulation showed that sensory prediction was also updated to predict perturbed auditory feedback (i.e., higher prediction in F1). Thus, prediction errors, the difference between lower perturbed feedback in F1 and higher prediction in F1, became minimized throughout adaptation, eventually becoming a small amount that could no longer induce adaptation.

Empirical evidence for the idea that minimal prediction errors may result in halting adaptation can be found in head movement adaptation of rhesus monkeys (Brooks et al., 2015). In the study, cerebellar neuron activities to the voluntary head movement became more suppressed (compared to passive movement) as adaptation plateaued. Critically, the authors argued that the neural response becoming more suppressed (or less "sensitive") throughout learning demonstrates that sensory prediction was being rapidly updated to predict unexpected (perturbed) sensory feedback.

Consistent with the previous finding, in the current study the late learning phase SIS increased (i.e., minimal prediction errors) in multiple participants who also showed plateaued adaptation in the phase (i.e., no additional learning). Furthermore, the observation that adaptation plateaued even though adaptation was largely incomplete (i.e., 14.88% of the perturbation size) can be best explained by the idea that sensory forward model updates (i.e., prediction updates) may have occurred throughout adaptation, minimalizing prediction errors. Thus, our findings add further support to the notion that incomplete adaptation may result from not only the motor output changes but also sensory prediction updates, which together minimize prediction errors.

What does SIS reflect?

SIS is typically viewed as a measure that reflects prediction errors given that it is reduced upon unexpected auditory feedback (e.g., pitch perturbation, alien voice). This view is also shared by other studies examining suppression of motor-evoked auditory responses (i.e., finger pressing a button), which is also reduced or absent in deviant (i.e., unpredicted) sounds (Knolle et al., 2013). In contrast to this view, a previous study from our laboratory argued that the SIS response may instead reflect target errors, discrepancies between an intended auditory target with auditory feedback (Niziolek et al., 2013). The study found that production variability reduced SIS. The study looked at formants at vowel onset and found that the greater the onset formants deviated from the median formants, the more SIS was reduced. Additionally, this reduction in SIS correlated with the amount of subsequent within-utterance formant change that reduced variance from the median as the utterance progressed ("centering"). Under the assumption that the median formants are close to the intended auditory target (i.e., an ideal production), the study argued that SIS reflects target errors.

However, our finding that SIS increased in 8 participants during the late learning phase cannot be easily explained by this account. Due to the SIS recovery, their late learning phase SIS response, which did not differ from their baseline SIS response, would be interpreted as minimal or no target errors according to the target error explanation for SIS. Nonetheless, these participants compensated for only 14.88% of the perturbation on average, presumably leaving a considerable discrepancy between any fixed auditory target and auditory feedback. Although previous studies have reported perceptual boundaries shifting towards the direction of perturbation during adaptation which may reduce target errors (Lametti et al., 2014; Shiller et al., 2009), it has also been suggested that auditory targets, as opposed to perceptual boundaries, do not change throughout adaptation (K. S. Kim & Max, 2021). In fact, a recent study has demonstrated that playing back the median production (i.e., the assumed auditory target) to participants throughout adaptation did not affect learning (LeBovidge et al., 2020), raising questions about whether auditory targets change during adaptation.

On the other hand, if SIS indeed reflects prediction errors rather than target errors, this view offers a different interpretation of Niziolek et al. (2013). According to the view, reduced SIS in productions with greater deviations from the median production may have resulted from large signal-dependent noise that stemmed from both the lower neural and muscular motor systems (Harris & Wolpert, 1998; Jones et al., 2002). Because such noise cannot be predicted by cortical areas, observed auditory feedback would not match auditory prediction, leading to large auditory prediction errors. Hence, it is plausible that the reduced SIS found in those productions reflects larger prediction errors. This view would also imply that centering (i.e., subsequent within-utterance formant change) minimized prediction errors, rather than target errors.

Neural correlates of auditory prediction errors

In the current study, we estimated auditory prediction errors from activities in the auditory cortex, but a large body of evidence suggests that the cerebellum may be a neural substrate for forward models that generate sensory predictions (e.g., Blakemore et al., 1999, 2001; Imamizu & Kawato, 2012; Kawato et al., 2003; Pasalar et al., 2006; Shadmehr, 2020; Shadmehr & Krakauer, 2008; Skipper & Lametti, 2021; Therrien & Bastian, 2019; Wolpert et al., 1998). Studies have also documented evidence that the cerebellum may also compute sensory prediction errors (e.g., Blakemore et al., 2001; Brooks et al., 2015; Cullen & Brooks, 2015). On the other hand, it has also been hypothesized that the cerebellum may work in concert with cortical areas to generate sensory prediction mechanisms and prediction errors (Blakemore & Sirigu, 2003; Haar & Donchin, 2020). In fact, the cerebellum is known to modulate activities in different cortical areas during active movements (e.g., the somatosensory cortex, Blakemore et al., 1999). Additionally, the cerebellum's projection to the posterior parietal cortex (Clower et al., 2001) has been implicated for generating sensory prediction(e.g., Della-Maggiore et al., 2004; Desmurget & Grafton, 2000; also see Blakemore & Sirigu, 2003 for a detailed review).

Is it possible that the cerebellum works in concert with the auditory cortex to compute auditory prediction errors? The cerebellum is certainly known for its involvement in auditory processing (e.g., Aitkin & Boyd, 1975, 1978; Ohyama et al., 2003) including speech perception (Ackermann et al., 2007; Mathiak et al., 2002; Schwartze & Kotz, 2016; Skipper & Lametti, 2021). It is also known that the cerebellum projects to the medial geniculate body (MGB), and the resulting inhibition and/or potentiation of MGB neurons may lead to rapid plasticity of response fields of the primary auditory cortex, modulating auditory inputs (e.g., McLachlan & Wilson, 2017; Weinberger, 2011). Such rapid plasticity of the response fields may prepare the primary auditory cortex for discriminating different sounds (David et al., 2012), a function that may be involved in computing auditory prediction errors. Indeed, both the right cerebellar areas and bilateral superior temporal cortex were found to be active during speech response to unexpected auditory error (i.e., under the presence of auditory prediction errors, Tourville et al., 2008).

Although studies have suggested that there is no direct projection from the primary auditory area to the cerebellum in primates (e.g., Schmahmann & Pandya, 1991) and mice (e.g., Henschke & Pakan, 2020), others reported auditory fibers projecting from the association areas and superior temporal gyrus to the cerebellum in primates (e.g., Brodal, 1979). In addition, it is also known that cortical auditory areas project to the cerebellar hemisphere through the cerebro–pontine pathways in some mammals including humans (e.g., Glickstein, 1997; Pastor et al., 2008). Taken together, although the exact neural correlates of auditory prediction errors remain largely unclear, it is possible that they are also computed through pathways/loops that involve multiple cortical and cerebellar areas.

It is also noteworthy that the baseline SIS activities were found to be most pronounced in the left auditory cortex, in line with the notion that the left hemisphere is dominant in speech and language perception (Curio et al., 2000; Houde et al., 2002). We also found SIS reduction only in the left auditory cortex, in line with a previous study that found prediction-related SIS effect only in the left hemisphere (Niziolek et al., 2013). One discrepancy in our finding from the previous study is that we did not find a significant SIS effect in the right hemisphere even during the baseline phase (see Supplemental Information 2). Given that the right hemisphere SIS is known to be highly variable across tasks and individuals (K. X. Kim et al., 2023), the discrepancy may have been due to the sampling issue.

Methods

Subjects

Across the two experiments (adaptation and control, see below), twenty-seven adult subjects who were 18 years of age or older without any speech, language, and hearing disorders were recruited. All subjects were native speakers of American English with no known communication, neurological, or psychological disorders. In addition, they passed pure-tone hearing thresholds of ≤ 20 dB HL for the octave frequencies between 500 and 4,000 Hz.

Because four subjects participated in both experiments 1-2 months apart, we obtained data from 31 sessions (21 sessions for adaptation and 9 sessions for control). Two of the four subjects participated in the adaptation experiment first. In the adaptation experiment, 7 subjects were excluded from analyses for various reasons. One subject's source could not be reliably localized, and three subjects could not finish the task due to fatigue. Two subjects showed "following" non-adaptive behavior and one subject had atypical SIS response in the baseline, (SIS < –5 z). Here, we report adaptation experiment results from 14 subjects (mean age = 31.5, SD = 9.8 years old, 8 females). For the control experiment, 1 subject was excluded because the subject's MRI could not be obtained, leaving 8 subjects for data analyses (mean age = 34.4, SD = 8.3 years old, 3 females).

Tasks

Adaptation

During MEG data collection of the first two sessions, subjects were asked to read "Ed," "end," or "ebb" (60 trial blocks for 3 different words = 180 total trials) that appeared on the screen. During these speaking sessions, subjects heard their speech with the first formant frequency (Formant 1 or F1) shifted upward for some trials (trial block 21 to 50, see below), which made their speech to sound more like "Add," "And," and "Ab," respectively. The auditory perturbation, 150 Hz upshift, was applied through Feedback Utility for Speech Processing (FUSP, Kothare et al., 2020) and the total feedback latency (i.e., hardware + software, K. S. Kim et al., 2020) was estimated to be about 19 ms.

During the speaking sessions, the first 20 trial blocks (i.e., baseline) had no perturbation, while blocks 21 through 50 had a 150 Hz up-shift perturbation in the auditory feedback. We categorized the first 15 trial blocks of the perturbed trials (21 – 45) as the early learning phase and the second 15 trial blocks (36-50) as the late learning phase. In the passive listening condition, subjects heard the same auditory feedback that they received during the speaking condition (including the perturbed sounds) through the earphones. With a mean interstimulus interval of 3s and short breaks (roughly 20 seconds) every 30 utterances, the duration of each condition was approximately 10 – 12 minutes. Given that the adaptation task (speak) was repeated, we also checked whether there was any savings effect and found that there was no consistent effect of repeating adaptation (see Supplemental Information 1).

Control

We also designed a control experiment in which we applied 0 Hz perturbation (instead of 150 Hz perturbation) during early and late "learning" phases. All other details of the task remained identical to the adaptation experiment.

MRI

On a separate day, subjects also underwent an MRI scan, where a high-resolution T1-weighted anatomical MRI was acquired in each participant for source reconstruction.

MEG acquisition

Subjects were placed in a 275-channel, whole-head biomagnetometer system (Omega 2000, CTF, Coquitlam, BC, Canada; sampling rate 1200 Hz; acquisition filtering 0.001-300 Hz) for a total of four sessions (two speaking and two listening sessions). Subjects heard auditory feedback (or recorded auditory feedback during listening condition) through ER-3A ear-insert earphones (Etymotic Research, Inc., Elk Grove Village, IL) and a passive fiber optic microphone (Phone-Or Ltd., Or-Yehuda, Israel) was placed about an inch in front of their mouths to record speech responses. All stimulus and response events were integrated in real time with MEG timeseries via analog-to-digital input to the imaging acquisition software.

Each subject lay supine with their head supported inside the helmet along the center of the sensor array. Three localizer coils affixed to the nasion, left peri-auricular, and right peri-auricular points determined head positioning relative to the sensor array both before and after each block of trials. We ensured that subjects' head movements were smaller than 5 mm in every session. Co-registration of MEG data to each individual's MRI image was performed using the CTF software suite (MISL Ltd., Coquitlam, BC, Canada; ctfmeg.com; version 5.2.1) by aligning the localizer coil locations to the corresponding fiducial points on the individual's MRI. MRI images were exported to Analyze format and spatially normalized to the standard T1 Montreal Neurological Institute (MNI) template via Statistical Parametric Mapping (SPM8, Wellcome Trust Centre for Neuroimaging, London, UK).

Data extraction and analyses

First formant frequency (F1)

The first formant frequency (F1) from each speech production was extracted through a custom MATLAB software, Wave Viewer (Raharjo et al., 2021). We then extracted F1 from the vowel midpoint (40% to 60% into the vowel) and averaged it for each utterance. In case of missing trials, we replaced the data point by using an interpolation method using four nearest neighboring trials as described in Kitchen et al. (2022). We replaced about 2.96% and 2.88% of the data for the adaptation and control experiments respectively. We normalized the data by subtracting the baseline F1 from the data (i.e., baseline = 6th to 20th trial blocks). The amount of learning in each phase was assessed by averaging the last 5 trial blocks (31st to 35th blocks for early learning and 46th to 50th blocks for late learning).

Speaking-induced suppression

We first corrected distant magnetic field disturbances by calculating a synthetic third-order gradiometer, detrended using a DC offset across whole trials, and then filtered (4th order

Butterworth, bandpass 4 to 40 Hz) sensor data. One subject's data in which we found additional noise caused by dental artifact through visual inspection was denoised using a dual signal subspace projection (DSSP, Cai, Kang, et al., 2019; Cai, Xu, et al., 2019). After pre-processing sensor data, separate datasets were created with trials during baseline, early learning, and late learning phases for speak and listen conditions. In these datasets, trials exceeding a 2 pT threshhold at any timepoint were rejected. In two subjects' data, three channels were removed prior to threshold-based artifact rejection. The data was then averaged across all remaining channels. For the adaptation experiment, 3.97% of the speak session trials and 3.78% of the listen session trials were removed. For the control experiment, 7.18% and 7.55% of the trials were removed for speak and listen sessions, respectively.

For each subject, a single-sphere head model was derived from the individual's co-registered T1 structural MRI using the CTF software suite (MISL Ltd., Coquitlam, BC, Canada; ctfmeg.com; version 5.2.1). Using the Champagne algorithm (Owen et al., 2012) and a lead field of 8mm resolution on the baseline listen data, we generated whole-brain evoked activity between 75 ms and 130 ms (after the auditory feedback onset), and determined the MNI coordinate with the most pronounced M100 response in the left and right auditory areas (i.e., the highest amplitude) for each subject. Although we only report the results from the left auditory area in the main text, the results for the right hemisphere can be found in the Supplemental Information 2. The median MNI coordinate across both adaptation and control experiments were [x = –56, y = –24, z = 0] and [x = 48, y = –16, z = 8] for the left and right auditory areas respectively. We then used a Bayesian adaptive beamformer (Cai et al., 2023) to extract time-series source activity focused on the obtained MNI coordinate across all phases (i.e., baseline, early, and late). From the final time-series z-scored data, we measured M100 peak by finding the maximum value between 75 – 130 ms after the auditory signal. We then computed the difference between the listen and speak sessions to determine SIS:

$$SIS = M100_{listen} - M100_{speak}$$

Statistical analysis

A linear mixed effects model was constructed for SIS with the different adaptation phases as fixed effects and subjects as a random effect using *lme4* package in R (Bates et al., 2015). The Tukey test was used for post-hoc pairwise comparisons from the *emmeans* package in R (Lenth, 2022). A Pearson's correlation tested to examine relationships between the amount of adaptation and the SIS amplitudes.

References

Abur, D., Lester-Smith, R. A., Daliri, A., Lupiani, A. A., Guenther, F. H., & Stepp, C. E. (2018). Sensorimotor adaptation of voice fundamental frequency in Parkinson's disease. *PloS One*, *13*(1), e0191839. https://doi.org/10.1371/journal.pone.0191839

Ackermann, H., Mathiak, K., & Riecker, A. (2007). The contribution of the cerebellum to speech production and speech perception: Clinical and functional imaging data. *Cerebellum (London, England)*, *6*(3), 202–213. https://doi.org/10.1080/14734220701266742

Aitkin, L. M., & Boyd, J. (1975). Responses of single units in cerebellar vermis of the cat to monaural and binaural stimuli. *Journal of Neurophysiology*, *38*(2), 418–429. https://doi.org/10.1152/jn.1975.38.2.418

Aitkin, L. M., & Boyd, J. (1978). Acoustic input to the lateral pontine nuclei. *Hearing Research*, *1*(1), 67–77. https://doi.org/10.1016/0378-5955(78)90010-2

Albert, S. T., Jang, J., Modchalingam, S., 't Hart, B. M., Henriques, D., Lerner, G., Della-Maggiore, V., Haith, A. M., Krakauer, J. W., & Shadmehr, R. (2022). Competition between parallel sensorimotor learning systems. *eLife*, *11*, e65361. https://doi.org/10.7554/eLife.65361

Albert, S. T., Jang, J., Sheahan, H. R., Teunissen, L., Vandevoorde, K., Herzfeld, D. J., & Shadmehr, R. (2021). An implicit memory of errors limits human sensorimotor adaptation. *Nature Human Behaviour*, *5*(7), 920–934. https://doi.org/10.1038/s41562-020-01036-x

Alemi, R., Lehmann, A., & Deroche, M. L. D. (2021). Changes in Spoken and Sung Productions Following Adaptation to Pitch-shifted Auditory Feedback. *Journal of Voice: Official*

Journal of the Voice Foundation, S0892-1997(21)00079-5. https://doi.org/10.1016/j.jvoice.2021.02.016

Avraham, G., Morehead, J. R., Kim, H. E., & Ivry, R. B. (2021). Reexposure to a sensorimotor perturbation produces opposite effects on explicit and implicit learning processes. *PLoS Biology*, *19*(3), e3001147. https://doi.org/10.1371/journal.pbio.3001147

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

Behroozmand, R., & Larson, C. R. (2011). Error-dependent modulation of speech-induced auditory suppression for pitch-shifted voice feedback. *BMC Neuroscience*, *12*, 54. https://doi.org/10.1186/1471-2202-12-54

Blakemore, S. J., Frith, C. D., & Wolpert, D. M. (2001). The cerebellum is involved in predicting the sensory consequences of action. *Neuroreport*, *12*(9), 1879–1884. https://doi.org/10.1097/00001756-200107030-00023

Blakemore, S. J., Wolpert, D. M., & Frith, C. D. (1999). The cerebellum contributes to somatosensory cortical activity during self-produced tactile stimulation. *NeuroImage*, *10*(4), 448–459. https://doi.org/10.1006/nimg.1999.0478

Blakemore, S.-J., & Sirigu, A. (2003). Action prediction in the cerebellum and in the parietal lobe. *Experimental Brain Research*, *153*(2), 239–245. https://doi.org/10.1007/s00221-003-1597-z

Brodal, P. (1979). The pontocerebellar projection in the rhesus monkey: An experimental study with retrograde axonal transport of horseradish peroxidase. *Neuroscience*, *4*(2), 193–208. https://doi.org/10.1016/0306-4522(79)90082-4

Brooks, J. X., Carriot, J., & Cullen, K. E. (2015). Learning to expect the unexpected: Rapid updating in primate cerebellum during voluntary self-motion. *Nature Neuroscience*, *18*(9), 1310–1317. https://doi.org/10.1038/nn.4077

Cai, C., Kang, H., Kirsch, H. E., Mizuiri, D., Chen, J., Bhutada, A., Sekihara, K., & Nagarajan, S. S. (2019). Comparison of DSSP and tSSS algorithms for removing artifacts from vagus nerve stimulators in magnetoencephalography data. *Journal of Neural Engineering*, *16*(6), 066045. https://doi.org/10.1088/1741-2552/ab4065

Cai, C., Long, Y., Ghosh, S., Hashemi, A., Gao, Y., Diwakar, M., Haufe, S., Sekihara, K., Wu, W., & Nagarajan, S. S. (2023). Bayesian Adaptive Beamformer for Robust Electromagnetic Brain Imaging of Correlated Sources in High Spatial Resolution. *IEEE Transactions on Medical Imaging*, *42*(9), 2502–2512. https://doi.org/10.1109/TMI.2023.3256963

Cai, C., Xu, J., Velmurugan, J., Knowlton, R., Sekihara, K., Nagarajan, S. S., & Kirsch, H. (2019). Evaluation of a dual signal subspace projection algorithm in magnetoencephalographic recordings from patients with intractable epilepsy and vagus nerve stimulators. *NeuroImage*, *188*, 161–170. https://doi.org/10.1016/j.neuroimage.2018.11.025

Chang, E. F., Niziolek, C. A., Knight, R. T., Nagarajan, S. S., & Houde, J. F. (2013). Human cortical sensorimotor network underlying feedback control of vocal pitch. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(7), 2653–2658. https://doi.org/10.1073/pnas.1216827110

Clower, D. M., West, R. A., Lynch, J. C., & Strick, P. L. (2001). The inferior parietal lobule is the target of output from the superior colliculus, hippocampus, and cerebellum. *The*

Journal of Neuroscience: The Official Journal of the Society for Neuroscience, *21*(16), 6283–6291. https://doi.org/10.1523/JNEUROSCI.21-16-06283.2001

Cullen, K. E., & Brooks, J. X. (2015). Neural correlates of sensory prediction errors in monkeys: Evidence for internal models of voluntary self-motion in the cerebellum. *Cerebellum (London, England)*, *14*(1), 31–34. https://doi.org/10.1007/s12311-014-0608-x

Curio, G., Neuloh, G., Numminen, J., Jousmäki, V., & Hari, R. (2000). Speaking modifies voice-evoked activity in the human auditory cortex. *Human Brain Mapping*, *9*(4), 183–191. https://doi.org/10.1002/(SICI)1097-0193(200004)9:4<183::AID-HBM1>3.0.CO;2-Z

Daliri, A., & Dittman, J. (2019). Successful auditory motor adaptation requires task-relevant auditory errors. *Journal of Neurophysiology*, *122*(2), 552–562. https://doi.org/10.1152/jn.00662.2018

David, S. V., Fritz, J. B., & Shamma, S. A. (2012). Task reward structure shapes rapid receptive field plasticity in auditory cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(6), 2144–2149. https://doi.org/10.1073/pnas.1117717109

Della-Maggiore, V., Malfait, N., Ostry, D. J., & Paus, T. (2004). Stimulation of the posterior parietal cortex interferes with arm trajectory adjustments during the learning of new dynamics. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *24*(44), 9971–9976. https://doi.org/10.1523/JNEUROSCI.2833-04.2004

Desmurget, M., & Grafton, S. (2000). Forward modeling allows feedback control for fast reaching movements. *Trends in Cognitive Sciences*, *4*(11), 423–431. https://doi.org/10.1016/s1364-6613(00)01537-0

Eliades, S. J., & Tsunada, J. (2018). Auditory cortical activity drives feedback-dependent vocal control in marmosets. *Nature Communications*, *9*(1), 2540. https://doi.org/10.1038/s41467-018-04961-8

Feng, Y., Gracco, V. L., & Max, L. (2011). Integration of auditory and somatosensory error signals in the neural control of speech movements. *Journal of Neurophysiology*, *106*(2), 667–679. https://doi.org/10.1152/jn.00638.2010

Glickstein, M. (1997). Mossy-fibre sensory input to the cerebellum. *Progress in Brain Research*, *114*, 251–259. https://doi.org/10.1016/s0079-6123(08)63368-3

Haar, S., & Donchin, O. (2020). A Revised Computational Neuroanatomy for Motor Control. *Journal of Cognitive Neuroscience*, *32*(10), 1823–1836. https://doi.org/10.1162/jocn_a_01602

Haith, A. M., & Krakauer, J. W. (2013). Model-based and model-free mechanisms of human motor learning. *Advances in Experimental Medicine and Biology*, *782*, 1–21. https://doi.org/10.1007/978-1-4614-5465-6_1

Harris, C. M., & Wolpert, D. M. (1998). Signal-dependent noise determines motor planning. *Nature*, *394*(6695), 780–784. https://doi.org/10.1038/29528

Heinks-Maldonado, T. H., Mathalon, D. H., Gray, M., & Ford, J. M. (2005). Fine-tuning of auditory cortex during speech production. *Psychophysiology*, *42*(2), 180–190. https://doi.org/10.1111/j.1469-8986.2005.00272.x

Heinks-Maldonado, T. H., Nagarajan, S. S., & Houde, J. F. (2006). Magnetoencephalographic evidence for a precise forward model in speech production. *Neuroreport*, *17*(13), 1375–1379. https://doi.org/10.1097/01.wnr.0000233102.43526.e9

Henschke, J. U., & Pakan, J. M. (2020). Disynaptic cerebrocerebellar pathways originating from multiple functionally distinct cortical areas. *eLife*, *9*, e59148. https://doi.org/10.7554/eLife.59148

Houde, J. F., Nagarajan, S. S., Sekihara, K., & Merzenich, M. M. (2002). Modulation of the auditory cortex during speech: An MEG study. *Journal of Cognitive Neuroscience*, *14*(8), 1125–1138. https://doi.org/10.1162/089892902760807140

Huang, V. S., Haith, A., Mazzoni, P., & Krakauer, J. W. (2011). Rethinking motor learning and savings in adaptation paradigms: Model-free memory for successful actions combines with internal models. *Neuron*, *70*(4), 787–801. https://doi.org/10.1016/j.neuron.2011.04.012

Imamizu, H., & Kawato, M. (2012). Cerebellar internal models: Implications for the dexterous use of tools. *Cerebellum (London, England)*, *11*(2), 325–335. https://doi.org/10.1007/s12311-010-0241-2

Jones, K. E., Hamilton, A. F., & Wolpert, D. M. (2002). Sources of signal-dependent noise during isometric force production. *Journal of Neurophysiology*, *88*(3), 1533–1544. https://doi.org/10.1152/jn.2002.88.3.1533

Kawato, M., Kuroda, T., Imamizu, H., Nakano, E., Miyauchi, S., & Yoshioka, T. (2003). Internal forward models in the cerebellum: fMRI study on grip force and load force coupling. *Progress in Brain Research*, *142*, 171–188. https://doi.org/10.1016/S0079-6123(03)42013-X

Keough, D., Hawco, C., & Jones, J. A. (2013). Auditory-motor adaptation to frequency-altered auditory feedback occurs when participants ignore feedback. *BMC Neuroscience*, *14*, 25. https://doi.org/10.1186/1471-2202-14-25

Kim, H. E., Parvin, D. E., & Ivry, R. B. (2019). The influence of task outcome on implicit motor learning. *eLife*, *8*, e39882. https://doi.org/10.7554/eLife.39882

Kim, K. S., Gaines, J. L., Parrell, B., Ramanarayanan, V., Nagarajan, S. S., & Houde, J. F. (2023). Mechanisms of sensorimotor adaptation in a hierarchical state feedback control model of speech. *PLoS Computational Biology*, *19*(7), e1011244. https://doi.org/10.1371/journal.pcbi.1011244

Kim, K. S., & Max, L. (2021). Speech auditory-motor adaptation to formant-shifted feedback lacks an explicit component: Reduced adaptation in adults who stutter reflects limitations in implicit sensorimotor learning. *The European Journal of Neuroscience*, *53*(9), 3093–3108. https://doi.org/10.1111/ejn.15175

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: JSLHR*, *63*(8), 2522–2534. https://doi.org/10.1044/2020_JSLHR-19-00419

Kim, K. X., Dale, C. L., Ranasinghe, K. G., Kothare, H., Beagle, A. J., Lerner, H., Mizuiri, D., Gorno-Tempini, M. L., Vossel, K., Nagarajan, S. S., & Houde, J. F. (2023). Impaired Speaking-Induced Suppression in Alzheimer's Disease. *eNeuro*, *10*(6), ENEURO.0056-23.2023. https://doi.org/10.1523/ENEURO.0056-23.2023

Kitchen, N. M., Kim, K. S., Wang, P. Z., Hermosillo, R. J., & Max, L. (2022). Individual sensorimotor adaptation characteristics are independent across orofacial speech movements and limb reaching movements. *Journal of Neurophysiology*. https://doi.org/10.1152/jn.00167.2022

Knolle, F., Schröger, E., & Kotz, S. A. (2013). Cerebellar contribution to the prediction of self-initiated sounds. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *49*(9), 2449–2461. https://doi.org/10.1016/j.cortex.2012.12.012

Kothare, H., Raharjo, I., Ramanarayanan, V., Ranasinghe, K., Parrell, B., Johnson, K., Houde, J. F., & Nagarajan, S. S. (2020). Sensorimotor adaptation of speech depends on the direction of auditory feedback alteration. *The Journal of the Acoustical Society of America*, *148*(6), 3682. https://doi.org/10.1121/10.0002876

Lametti, D. R., Krol, S. A., Shiller, D. M., & Ostry, D. J. (2014). Brief periods of auditory perceptual training can determine the sensory targets of speech motor learning. *Psychological Science*, *25*(7), 1325–1336. https://doi.org/10.1177/0956797614529978

Lametti, D. R., Quek, M. Y. M., Prescott, C. B., Brittain, J.-S., & Watkins, K. E. (2020). The perils of learning to move while speaking: One-sided interference between speech and visuomotor adaptation. *Psychonomic Bulletin & Review*, *27*(3), 544–552. https://doi.org/10.3758/s13423-020-01725-8

LeBovidge, E., Li, C., & Max, L. (2020). *Toward understanding the limiting factors in speech auditory-motor adaptation: A new look at perceptual targets*. 12th International Seminar on Speech Production, New Haven, CT. https://issp2020.yale.edu/

Lenth, R. V. (2022). *emmeans: Estimated Marginal Means, aka Least-Squares Means* [Computer software]. https://CRAN.R-project.org/package=emmeans

Leow, L.-A., Marinovic, W., de Rugy, A., & Carroll, T. J. (2018). Task errors contribute to implicit aftereffects in sensorimotor adaptation. *The European Journal of Neuroscience*, *48*(11), 3397–3409. https://doi.org/10.1111/ejn.14213

Leow, L.-A., Marinovic, W., de Rugy, A., & Carroll, T. J. (2020). Task Errors Drive Memories That Improve Sensorimotor Adaptation. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *40*(15), 3075–3088. https://doi.org/10.1523/JNEUROSCI.1506-19.2020

Lester-Smith, R. A., Daliri, A., Enos, N., Abur, D., Lupiani, A. A., Letcher, S., & Stepp, C. E. (2020). The Relation of Articulatory and Vocal Auditory-Motor Control in Typical Speakers. *Journal of Speech, Language, and Hearing Research: JSLHR*, *63*(11), 3628–3642. https://doi.org/10.1044/2020_JSLHR-20-00192

Mathiak, K., Hertrich, I., Grodd, W., & Ackermann, H. (2002). Cerebellum and speech perception: A functional magnetic resonance imaging study. *Journal of Cognitive Neuroscience*, *14*(6), 902–912. https://doi.org/10.1162/089892902760191126

Max, L., & Maffett, D. G. (2015). Feedback delays eliminate auditory-motor learning in speech production. *Neuroscience Letters*, *591*, 25–29. https://doi.org/10.1016/j.neulet.2015.02.012

Mazzoni, P., & Krakauer, J. W. (2006). An implicit plan overrides an explicit strategy during visuomotor adaptation. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *26*(14), 3642–3645. https://doi.org/10.1523/JNEUROSCI.5317-05.2006

McLachlan, N. M., & Wilson, S. J. (2017). The Contribution of Brainstem and Cerebellar Pathways to Auditory Recognition. *Frontiers in Psychology*, *8*, 265. https://doi.org/10.3389/fpsyg.2017.00265

Miyamoto, Y. R., Wang, S., & Smith, M. A. (2020). Implicit adaptation compensates for erratic explicit strategy in human motor learning. *Nature Neuroscience*, *23*(3), 443–455. https://doi.org/10.1038/s41593-020-0600-3

Morehead, J. R., & Xivry, J.-J. O. de. (2021). *A Synthesis of the Many Errors and Learning Processes of Visuomotor Adaptation* (p. 2021.03.14.435278). bioRxiv. https://doi.org/10.1101/2021.03.14.435278

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

Niziolek, C. A., Nagarajan, S. S., & Houde, J. F. (2013). What does motor efference copy represent? Evidence from speech production. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *33*(41), 16110–16116. https://doi.org/10.1523/JNEUROSCI.2137-13.2013

Ohyama, T., Nores, W. L., Murphy, M., & Mauk, M. D. (2003). What the cerebellum computes. *Trends in Neurosciences*, *26*(4), 222–227. https://doi.org/10.1016/S0166-2236(03)00054-7

Owen, J. P., Sekihara, K., & Nagarajan, S. S. (2012). Non-parametric statistical thresholding for sparse magnetoencephalography source reconstructions. *Frontiers in Neuroscience*, *6*, 186. https://doi.org/10.3389/fnins.2012.00186

Parrell, B., Ramanarayanan, V., Nagarajan, S., & Houde, J. (2019). The FACTS model of speech motor control: Fusing state estimation and task-based control. *PLoS Computational Biology*, *15*(9), e1007321. https://doi.org/10.1371/journal.pcbi.1007321

Pasalar, S., Roitman, A. V., Durfee, W. K., & Ebner, T. J. (2006). Force field effects on cerebellar Purkinje cell discharge with implications for internal models. *Nature Neuroscience*, *9*(11), 1404–1411. https://doi.org/10.1038/nn1783

Pastor, M. A., Vidaurre, C., Fernández-Seara, M. A., Villanueva, A., & Friston, K. J. (2008). Frequency-specific coupling in the cortico-cerebellar auditory system. *Journal of Neurophysiology*, *100*(4), 1699–1705. https://doi.org/10.1152/jn.01156.2007

Raharjo, I., Kothare, H., Nagarajan, S. S., & Houde, J. F. (2021). Speech compensation responses and sensorimotor adaptation to formant feedback perturbations. *The Journal of the Acoustical Society of America*, *149*(2), 1147. https://doi.org/10.1121/10.0003440

Sato, M., & Shiller, D. M. (2018). Auditory prediction during speaking and listening. *Brain and Language*, *187*, 92–103. https://doi.org/10.1016/j.bandl.2018.01.008

Schmahmann, J. D., & Pandya, D. N. (1991). Projections to the basis pontis from the superior temporal sulcus and superior temporal region in the rhesus monkey. *The Journal of Comparative Neurology*, *308*(2), 224–248. https://doi.org/10.1002/cne.903080209

Schwartze, M., & Kotz, S. A. (2016). Contributions of cerebellar event-based temporal processing and preparatory function to speech perception. *Brain and Language*, *161*, 28–32. https://doi.org/10.1016/j.bandl.2015.08.005

Shadmehr, R. (2020). Population coding in the cerebellum: A machine learning perspective. *Journal of Neurophysiology*, *124*(6), 2022–2051. https://doi.org/10.1152/jn.00449.2020

Shadmehr, R., & Krakauer, J. W. (2008). A computational neuroanatomy for motor control. *Experimental Brain Research*, *185*(3), 359–381. https://doi.org/10.1007/s00221-008-1280-5

Shiller, D. M., Mitsuya, T., & Max, L. (2020). Exposure to Auditory Feedback Delay while Speaking Induces Perceptual Habituation but does not Mitigate the Disruptive Effect of Delay on Speech Auditory-motor Learning. *Neuroscience*, *446*, 213–224. https://doi.org/10.1016/j.neuroscience.2020.07.041

Shiller, D. M., Sato, M., Gracco, V. L., & Baum, S. R. (2009). Perceptual recalibration of speech sounds following speech motor learning. *The Journal of the Acoustical Society of America*, *125*(2), 1103–1113. https://doi.org/10.1121/1.3058638

Skipper, J. I., & Lametti, D. R. (2021). Speech Perception under the Tent: A Domain-general Predictive Role for the Cerebellum. *Journal of Cognitive Neuroscience*, *33*(8), 1517–1534. https://doi.org/10.1162/jocn_a_01729

Therrien, A. S., & Bastian, A. J. (2019). The cerebellum as a movement sensor. *Neuroscience Letters*, *688*, 37–40. https://doi.org/10.1016/j.neulet.2018.06.055

Weinberger, N. M. (2011). The medial geniculate, not the amygdala, as the root of auditory fear conditioning. *Hearing Research*, *274*(1–2), 61–74. https://doi.org/10.1016/j.heares.2010.03.093

Wolpert, D. M., Miall, R. C., & Kawato, M. (1998). Internal models in the cerebellum. *Trends in Cognitive Sciences*, *2*(9), 338–347. https://doi.org/10.1016/s1364-6613(98)01221-2

Xia, M., Wang, J., & He, Y. (2013). BrainNet Viewer: A network visualization tool for human brain connectomics. *PloS One*, *8*(7), e68910. https://doi.org/10.1371/journal.pone.0068910

Supplemental Information 1: Repeated adaptation session

Unlike arm reaching adaptation studies showing that re-experiencing adaptation results in savings (e.g., Huang et al., 2011) or attenuation in case of implicit adaptation (Avraham et al., 2021), we did not find any significant changes on the repeated adaptation session (2nd session, see Fig. S1A). The un-normalized baseline phase (before normalizing the baseline phase to 0 Hz) showed no significant difference across the two sessions, $t(13) = -1.985$, $p = 0.069$ (Fig. S1B, left), though the p-value was close to 0.05 due to two individuals whose baseline was much lower in the 2nd session than the 1st session. Nonetheless, there was no clear trend on whether these individuals learned more or less in the 2nd session. Indeed, the normalized adaptation data clearly shows that both the early and late phase data in the repeated adaptation session was not different from the initial adaptation (Fig. S1B, right).

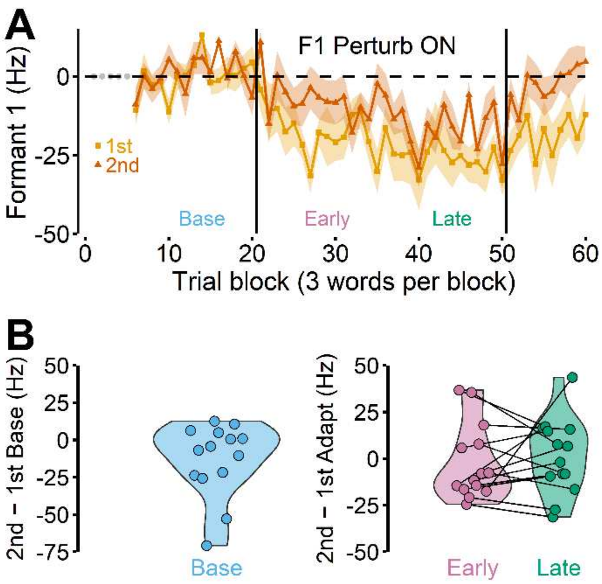

Fig. S1. A: Adaptation did not differ between the first and second sessions. B: The overall baseline also did not change in most participants (left). There were two participants whose baseline in the second session was reduced by more than 50 Hz, but as a group the baseline did not differ between the two sessions. Overall, participants as a group, there was no sign of savings or attenuation in the second session (right).

Supplemental Information 2: Right hemisphere data

Compared the left auditory cortex, the right hemisphere SIS activities were less pronounced. Multiple individuals did not show a clear SIS response in the right hemisphere even during the baseline phase in both the adaptation and control experiments (see Fig. S2). In addition, in the adaptation group, we did not observe a significant SIS reduction in adaptation phases in the right auditory cortex, F (2, 26) = 0.150, p = 0.862, in line with a previous study that found prediction-related SIS effect only in the left hemisphere (Niziolek et al., 2013). We also did not find any significant SIS reduction in the control group's right hemisphere activities, F (2, 16) = 0.854, p = 0.444. It should be noted that one participant's data in the adaptation group was excluded from analyses because source localization for the right hemisphere response was unreliable.

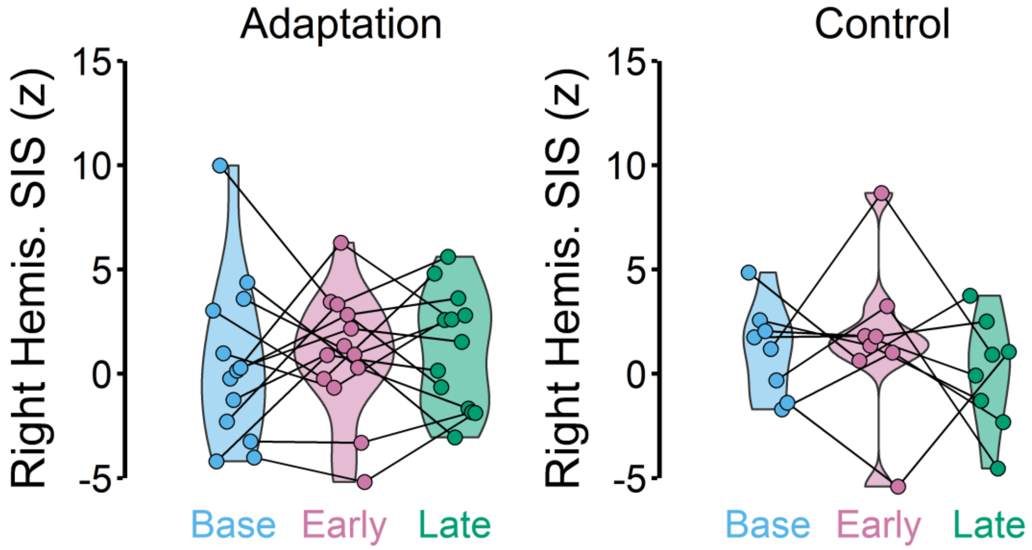

Fig. S2. We did not find a significant SIS effect during the baseline in the right hemisphere. In addition, we did not observe any significant SIS reduction during adaptation.

Supplemental Information 3: Outlier data

One participant's late learning phase SIS response was an outlier, which was imputed in our correlation analysis (see Results). One reason for keeping the data in our analysis was that the behavioral data also clearly indicated that the late learning response was not due to extreme points. As shown in Fig. S3, the participant clearly showed late learning.

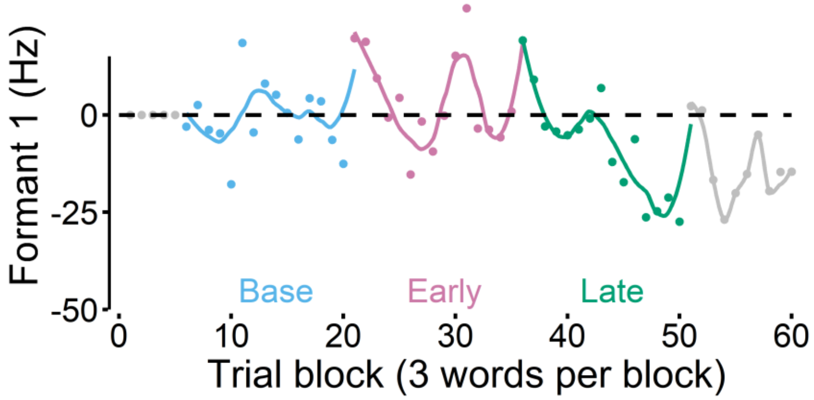

Fig. S3. A participant who had an outlier SIS response (i.e., a large SIS reduction) during the late learning phase also showed large adaptation in the same phase.