american physiological society
JNP JOURNAL OF NEUROPHYSIOLOGY

J Neurophysiol 133: 479–489, 2025.
First published December 31, 2024; doi:10.1152/jn.00337.2024

RESEARCH ARTICLE

Higher Neural Functions and Behavior

Probing sensorimotor memory through the human speech-audiomotor system

Nishant Rao[1,2] and **David J. Ostry**[1,2]

[1]*Yale Child Study Center, Yale School of Medicine, Yale University, New Haven, Connecticut, United States and* [2]*Department of Psychology, McGill University, Montreal, Quebec, Canada*

Abstract

Our knowledge of human sensorimotor learning and memory is predominantly based on the visuospatial workspace and limb movements. Humans also have a remarkable ability to produce and perceive speech sounds. We asked whether the human speech-auditory system could serve as a model to characterize the retention of sensorimotor memory in a workspace that is functionally independent of the visuospatial one. Using adaptation to altered auditory feedback, we investigated the durability of a newly acquired speech-acoustical memory (8- and 24-h delay), its sensitivity to the manner of acquisition (abrupt vs. gradual perturbation), and factors affecting memory retrieval. We observed extensive retention of learning (∼70%) but found no evidence for offline gains. The speech-acoustical memory was insensitive to the manner of its acquisition. To assess factors affecting memory retrieval, tests were first done in the absence of auditory feedback (with masking noise). Under these conditions, it appeared there was no memory for prior learning as if after an overnight delay, speakers had returned to their habitual speech production modes. However, when speech was reintroduced, resulting in speech error feedback, speakers returned immediately to their fully adapted state. This rapid switch shows that the two modes of speech production (adapted and habitual) can coexist in parallel in sensorimotor memory. The findings demonstrate extensive persistence of speech-acoustical memory and reveal context-specific memory retrieval processes in speech-motor learning. We conclude that the human speech-auditory system can be used to characterize sensorimotor memory in a workspace that is distinct from the visuospatial workspace.

NEW & NOTEWORTHY There is extensive retention of speech-motor learning. Two parallel modes exist in speech motor memory, one with access to everyday habitual speech and the other with access to newly learned speech-acoustical maps. The availability of speech error feedback triggers a switch between these two modes. Properties of sensorimotor memory in the human speech-auditory system are behaviorally similar to, but functionally independent of, their visuospatial counterparts.

auditory perception; contextual memory retrieval; implicit adaptation; speech memory retention; speech motor learning

INTRODUCTION

The benefits of sensorimotor learning are seen in the stabilization and retrieval of sensorimotor memories. Investigations of the characteristics of these memories have been predominantly conducted using limb movement paradigms (1–3). Although limb movement research has advanced our understanding of sensorimotor memories, much of this work has focused on the investigation of the visuospatial workspace, often using human and animal models (4–7). One of the unique characteristics of human behavior is our ability to learn, produce, and perceive sounds and form speech-acoustical memories (8–11). However, it remains unknown whether the human speech-auditory system could also be used to assess sensorimotor memory retention in the workspace that is functionally independent of the visuospatial workspace.

The formation of new speech-acoustical memories can be elicited by the introduction of auditory feedback shifts and associated speech error processing (9). For instance, perturbation of vowel formant frequency during word and sentence utterances leads to systematic changes in vocal output in directions opposite to the applied perturbation (12–15). Similar patterns of speech-motor learning have been shown to occur with perturbation of pitch, formant frequency, vowel duration, and intensity (16–19). The limb movement literature has found a dependence of visuospatial memory

Correspondence: D. J. Ostry (david.ostry@mcgill.ca).

Submitted 31 July 2024 / Revised 23 December 2024 / Accepted 24 December 2024

retention and retrieval on the nature of the task under investigation, the adaptation schedule, and the time elapsed following learning (20, 21). For instance, decay of motor memory has been found to be greater following learning in an abruptly changing environment compared with a gradually changing environment (22), which is consistent with other studies (23, 24). On the other hand, retention in a locomotor adaptation task was shown to be indifferent to the gradual or abrupt acquisition of these memories (25), corroborating evidence from other tasks involving reaching (26, 27).

Here, we used a vowel-formant-frequency perturbation task to ask a number of related questions in the context of speech, specifically, whether speech-acoustical memories are transient versus durable, whether these memories are sensitive to the manner of memory acquisition, whether speech-acoustical memories undergo offline gains, and whether the memory retrieval is dependent upon the availability of auditory feedback. Operationally, we assessed whether newly acquired speech acoustical memories are retained following a single session of acquisition and whether the retention (if any) is dependent upon the time elapsed since acquisition (8- vs. 24-h intervals) and/or the nature of the acquisition schedule (abrupt vs. gradual learning). We observed that the retention of speech-acoustical memory was robust and unaltered following either an 8- or 24-h delay and was no different for abrupt and gradual learning. Indeed, the new speech-acoustical memory was remarkably durable with little loss of information following 8- or 24-h delay, but there was little evidence of offline gains. In tests of the retrieval of the newly acquired memory, we found that speech memory retrieval was substantially degraded in the absence of speech-auditory feedback but was enabled almost immediately following the availability of speech and resulting speech feedback errors. This rapid switch shows that two modes of speech production (adapted and habitual) can coexist in parallel in sensorimotor memory. Overall, this study identifies properties of speech acoustical memory—it is highly durable, unaffected by the manner of memory acquisition, and its retrieval is context-specific and dependent on the availability of speech error feedback. This study establishes the human speech-auditory system as a model to probe, characterize, and compare speech-motor processing and memory with that of the visuospatial workspace and limb movement.

MATERIALS AND METHODS

Participants

A total of 77 young adults of either sex participated in this study. None of the participants reported speech or hearing disorders. Participants provided written informed consent, in accordance with the study procedures approved by the Institutional Review Boards of McGill University and Yale University. Data from four of these participants were excluded due to their follower-like behavior in the learning task, and one was removed due to technical issues during the session. Collectively, data from 72 participants (21 males; mean age ± SE: 23.94 ± 0.57) was considered for subsequent analyses. The proportion of female and male participants was similar across conditions.

Experimental Setup

Participants were comfortably seated and faced a computer monitor placed on a table in front of them (Fig. 1*A*). A unidirectional microphone (Sennheiser) was used to record their speech data on each trial. The speech signal was fed back in real-time through headphones (Beyerdynamic DT770M), either with no manipulation or following a specific experimental manipulation using Audapter (28) integrated with MATLAB. The microphone and headphones were interfaced with an audio mixer (MOTU/TASCAM) that enabled the separate control of gain to the headphones and microphone. The same feedback gains were used for all participants.

Speech Motor Learning Task

Vowels are acoustically organized in terms of vocal tract resonances known as formants, with the first and the second formant frequencies accounting for the most acoustical energy (11, 19). The primary procedure involved the real-time manipulation of the first formant frequency of spoken vowels in test words that were played back to participants through headphones (Fig. 1*B*). Specifically, participants read aloud consonant-vowel-consonant pseudowords ("bep," "dep," or "gep") displayed on a computer monitor, one at a time. The study consisted of two visits, each involving a combination of speech- and noise-feedback trials. During the first study visit (*visit 1*), participants performed 30 baseline trials with no formant perturbation (i.e., participants heard their own voice through the headphones). Three additional trials were interspersed in the baseline block during which speech-modulated noise feedback (referred to as noise feedback trials hereafter) was played to the participants through the headphones as they spoke and served to mask their own acoustical output. Providing noise feedback that was modulated by participants' speech in real time minimized the likelihood of a Lombard effect (speaking loudly with a noise background) during the noise feedback trials and provided a measure of vocal output with little opportunity for online correction. The baseline sequence was followed by a set of 210 trials during which the real-time first formant frequency (F1) was increased either gradually over 30 trials or abruptly by 30%. The gradual introduction of the F1 perturbation was uniform, starting with the first trial after the baseline block and ending with the maximum perturbation (i.e., 30% increase in F1) on the 30th trial of the learning block. In the trials between 30 and 210, the F1 perturbation was held at 30% throughout. *Visit 1* concluded with three noise feedback trials at the end of the learning block. Participants returned either ∼8 or ∼24 h later for a second visit (*visit 2*) which started with three noise feedback trials, followed by a set of 210 trials with a constant 30% F1 perturbation and concluded with three noise feedback trials at the end of the block (Fig. 1*C*). We chose to use pseudowords as opposed to real words to reduce the potential interference to the newly learned speech-acoustical mapping from routine speech outside of the laboratory between *visit 1* and *visit 2*. For all the participants, the auditory feedback level was increased so that the participants did not hear their own voice other than through the headphones. All participants were also instructed to speak softly and consistently

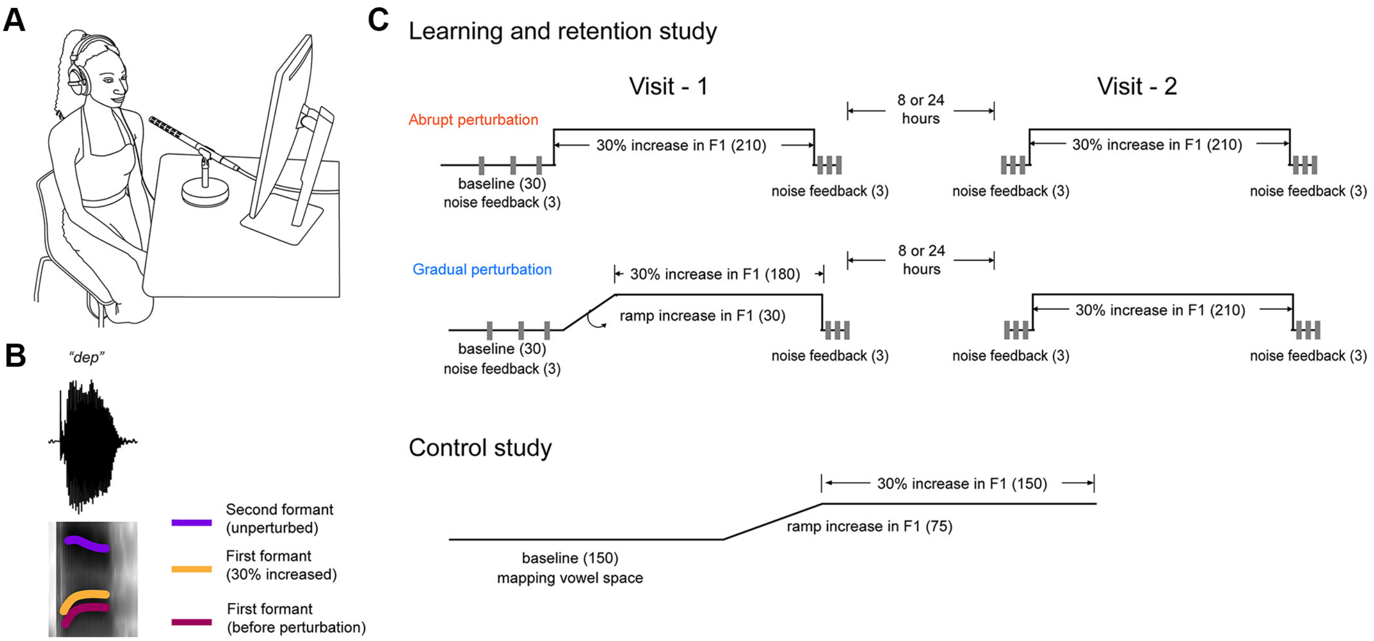

Figure 1. Experimental design. *A*: experimental setup with a participant seated in front of a computer monitor, which displayed the stimulus to be read aloud. The microphone and associated audio setup recorded the participant's speech and played it back through headphones. *B*: schematic of formant perturbation of a speech signal (time domain on *top*, spectrogram on the *bottom*). *C*: study design. Vertical gray stripes indicate noise feedback trials, the numbers in parentheses indicate number of trials for that condition.

throughout the experimental sessions and maintain a distance ∼15 cm from the microphone.

A control study was conducted to identify the region within the vowel space in which speech-motor learning stabilizes. The goal was to determine whether the response to a localized acoustical perturbation, that is to F1 alone, resulted in a change that was also in F1 alone, in effect producing a new vowel versus a derivative of an existing vowel. The control study used the vowel /Ɛ/ (as in "dep"), which was manipulated in the primary study, with respect to unperturbed neighboring vowels /ɪ/ (as in "dip") and /æ/ ("dap"). The control study involved reading aloud words displayed on a computer monitor, one at a time. The words were consonant-vowel-consonant sequences, "dap," "dep," or "dip," which were presented 50 times each in a pseudorandomized order during baseline. There was no perturbation during a baseline block. Trials following the baseline block only involved presentation and manipulation of the vowel /Ɛ/ in "dep"; the first formant frequency was gradually increased over 75 trials and then held constant for a further 150 trials (Fig. 1*C*). As in the primary study, the alteration involved changes only to the first formant frequency up to a maximum of 30%.

Data Analysis

Speech data were sampled at 48 kHz and then downsampled to 16 kHz to reduce processing time. Vowel boundaries were first identified using Montreal Forced Aligner (29) followed by manual inspection and correction when needed. First and second formant frequencies (F1 and F2) were extracted from the speech data on each trial using the Burg algorithm in Praat (30, 31).

A 40-ms sample was taken from the center of the vowel (20 ms before and after the vowel midpoint). To assess whether this window was early enough to tap into a feedforward component of vowel production, we conducted an analysis of vowel duration. The vowel duration values averaged 154 ms ± 3 ms (mean ± SE). There were no differences in duration across experimental conditions. Means ± SE were 156 ± 4 for abrupt and 153 ± 4 for gradual introduction of the perturbation. For 8- and 24-h retention intervals, values were 153 ± 5 and 156 ± 4 respectively. We also assessed vowel duration on a word-specific basis. On average the duration was 158 ± 5 for "bep," 153 ± 5 for "dep," and 150 ± 5 ms for "gep." Consequently, our formant assessment window ranged on average from 57 to 97 ms, which may limit any auditory feedback-based adjustments to vocal output (32, 33).

Following formant extraction, F1 and F2 values were averaged over the baseline trials with speech feedback, and percentage change from baseline was computed for all trials. We verified that there were no differences in baseline values across experimental conditions (abrupt vs. gradual, 8- vs. 24-h) nor for noise versus speech feedback trials. Subsequently, for visualization purposes, trials were binned over three consecutive trials without overlap to assess the time course of learning on each visit. A computation of retention as a percentage of learning was conducted by calculating the percentage change in F1 from baseline in the second bin after providing speech feedback on *visit 2*, normalized by percentage change in F1 in the last 10 bins with speech feedback on *visit 1* (learning asymptote). We focused specifically on the second bin for this test, as it provides a stable measure of retention, without transient effects attributable to the first bin immediately after reintroducing speech feedback. However, even when the first bin was used, reliable retention was observed. Data from participants whose asymptote in *visit 1* was in the direction of applied perturbation (so-called followers) were not considered specifically for this analysis (three individuals).

An analysis of context dependency compared data for utterances with masking noise feedback with that of speech acoustical feedback. The effect of speech feedback on learning and retention was assessed by the percentage change in F1 from baseline at the end of learning (last bin with speech feedback vs. last bin with noise feedback on *visit 1*), and at the start of *visit 2* (first bin with noise feedback vs. first bin with speech feedback on *visit 2*). It should be noted that for this analysis we used a common baseline for normalization rather than separately normalizing speech and noise feedback trials with their respective baseline. We did so because the difference in F1 frequency between speech and noise feedback trials at baseline was not significantly different ($F_{1,114} = 0.127$, $P = 0.722$). Moreover, there were only three noise feedback trials at baseline. Hence, normalizing noise trials throughout their own baseline may well add unnecessary variation to the data.

In all cases, ANOVA was used to test for differences between experimental conditions. Repeated-measures *t* tests were used to compare performance between noise feedback and speech feedback trials. One-sample *t* tests were used to assess whether the extent of learning and retention as assessed in noise and speech feedback was different from zero. Bonferroni corrections were used where applicable.

Control Study

Data sampling, scoring, and extraction of F1 and F2 values were conducted as in the primary study aforementioned. Following the formant extraction, F1 values were averaged over the baseline trials (“dep”) and the percentage change in F1 from baseline was computed for the learning trials (also for “dep”). We removed trials beyond 2.5 standard deviation of the mean for each word (such trials constituted <3% of the total trials). F1 and F2 frequencies were transformed into *z*-scores on a per subject basis using values for all vowels (50 trials each for “dap,” “dep,” and “dip” during baseline and the 50 trials at the end of learning for “dep”). A subsequent vector analysis was conducted to assess the magnitude and direction of the acoustical change following learning from each of the utterances under baseline conditions. This involved subtracting *z*-scores of “dep (baseline)” from “dap” and “dip,” and of “dep (end of learning)” from “dap” and “dip” for each participant. Also, the *z*-score of “dep (baseline)” was subtracted from that of “dep (end of learning)” to obtain the resultant shift for that vowel after learning for each participant.

RESULTS

The primary goal of this study was to assess the durability and robustness of newly acquired speech-acoustical memories in the context of speech-motor learning. The experimental conditions probed the retention of new learning. Separate conditions involved participants learning to compensate for a real-time 30% increase in their first formant (F1) frequency following either abrupt or gradual introduction of the auditory feedback alteration and returning for a second visit either 8- or 24-h later. Speech-modulated noise feedback trials were introduced during both visits to determine whether access to the newly acquired memory was contextually dependent on the presence of speech error feedback.

Robust Adaptation and Retention following Perturbation of First Formant Frequency

The results of the primary experiment are shown in Figs. 2 and 3. Figure 2*A* shows *visit 1* and *visit 2* data in terms of F1 values that were separated either by 8 h (Fig. 2*A*, *top*, $n = 28$) or 24 h (Fig. 2*A*, *bottom*, $n = 30$). Solid circles indicate binned data averaged over three consecutive utterances without overlap. Colors differentiate the introduction of the formant perturbation during the first visit as either abrupt (red; $n = 13$ for 8-h and $n = 15$ for 24-h conditions) or gradual (blue; $n = 15$ for 8-h and $n = 15$ for 24-h conditions). The vertical gray bars indicate trials with speech-shaped noise feedback (referred to as noise feedback subsequently). It is seen that participants showed consistent learning and retention following perturbation of the first formant frequency. The retention was unaltered by the passage of time or the manner of introduction of the perturbation. Moreover, asymptotic performance did not differ between the two visits (mean change in F1 from baseline ± SE was $-12.46 \pm 0.87\%$ and $-12.19 \pm 0.94\%$ for *visit 1* and *visit 2*, respectively).

Retention was assessed both in the presence of noise feedback and when acoustical feedback was reintroduced. We computed retention when acoustical feedback was present by normalizing percentage change in F1 from baseline at the start of *visit 2* (second bin after receiving speech feedback) by the asymptotic value at the end of learning. The results are shown in Fig. 2*B*. As can be seen, there was extensive retention of the newly learned speech-acoustical map and little difference in retention with the passage of time or the perturbation schedule.

A statistical analysis of retention in the presence of acoustical feedback was conducted to assess the effect of the interval following learning (8- and 24-h) and the perturbation pattern (abrupt and gradual). The extent of retention was reliably different than zero with abrupt perturbation in both 8- and 24-h conditions ($t_{11} = 4.782$, adjusted $P = 0.001$, and $t_{14} = 5.804$, adjusted $P < 0.001$, respectively). Similarly, retention was reliably different than zero with gradual perturbation in 8- and 24-h conditions ($t_{14} = 6.939$, adjusted $P < 0.001$, and $t_{12} = 11.378$, adjusted $P < 0.001$, respectively). Moreover, the observed retention was unaltered with the interval postlearning or the perturbation pattern (Fig. 2*B*, Retention interval: $F_{1,51} = 0.427$, $P = 0.516$; Perturbation pattern: $F_{1,51} = 0.155$, $P = 0.694$; Retention interval × Perturbation pattern: $F_{1,51} = 0.039$, $P = 0.843$).

Figure 2*C* shows a comparison of learning and retention as assessed with noise feedback versus speech feedback. It is seen that there is evidence of learning both with noise feedback and with speech feedback. Although the measure of learning is less with noise feedback, learning differed little for the 8- and 24-h conditions, or in the abrupt and gradual perturbation conditions. In contrast, there was little evidence of retention with noise feedback in *visit 2*, whereas retention was clearly evident when speech feedback, leading to speech errors, was made available (indicated by differences in the lighter vs. darker shades of the same color).

To assess differences in learning and retention with auditory and noise feedback (Fig. 2*C*), we compared last bin of noise feedback with last bin of auditory feedback during *visit 1*, and the first bin of noise feedback and the first bin of speech

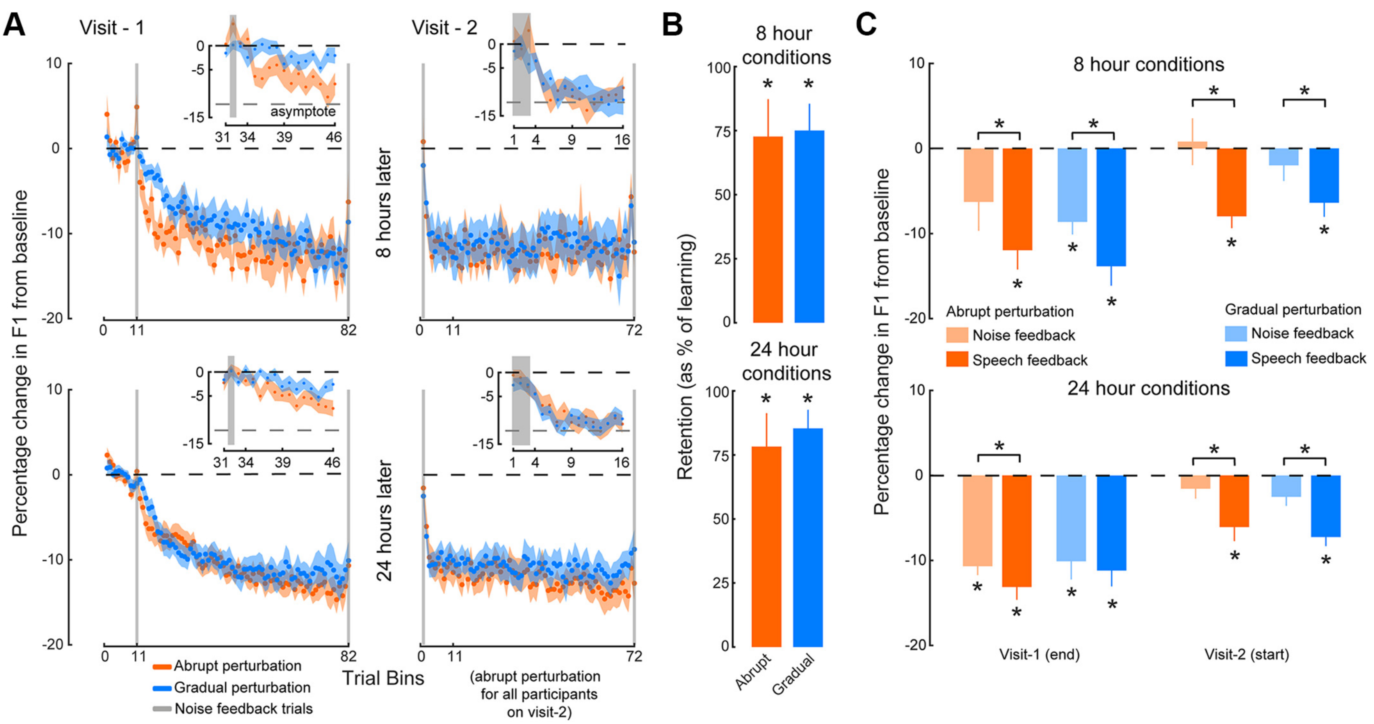

Figure 2. Extensive retention and context-based retrieval following speech motor learning. *A*: percentage change in the first formant (F1) from baseline during learning and retention. The horizontal dashed line indicates zero percentage change in F1 from baseline. Each solid circle gives the average over three consecutive trials (1 bin), the shaded region indicates standard error. Vertical gray bars indicate bins with speech-modulated noise feedback. During the baseline block, these trials were interspersed with nonperturbed trials, but are combined and shown at the end of baseline block for visualization purposes. *Insets* show initial trialwise data for *visit 1* and *visit 2*. *B*: retention as a percentage of learning (second bin after reintroducing speech feedback on *visit 2* normalized by average of last 10 bins with speech feedback on *visit 1*). *C*: comparison of learning (last bins) and retention (first bins) with speech feedback (darker shades) and those with speech-modulated noise feedback (lighter shades). *P values less than 0.05.

feedback during the *visit 2*. For purposes of this analysis, we pooled over retention interval and the perturbation pattern and conducted the analysis of noise feedback versus auditory feedback using ANOVA. Percentage change in F1 from baseline was found to be different for noise and speech feedback conditions during learning (Feedback type: $F_{1,114} = 5.678$, $P = 0.018$). Similarly, the percentage change in F1 from baseline also differed for speech versus noise feedback during retention (Feedback type: $F_{1,114} = 22.065$, $P < 0.001$).

Within the noise feedback condition, we assessed whether learning and retention were reliably different than zero. We found that the percentage change in F1 from baseline for noise feedback trials was reliably different than zero only after learning ($t_{12} = -1.770$, adjusted $P = 0.305$; $t_{14} = -9.919$, adjusted $P < 0.001$ for 8- and 24-h retention with abrupt training; $t_{14} = -5.625$, adjusted $P < 0.001$; $t_{12} = -4.574$, adjusted $P = 0.001$, for gradual training). The change relative to zero was not reliable in the noise feedback trials at the beginning of *visit 2* for 8- and 24-h with abrupt perturbation ($t_{12} = 0.282$, adjusted $P = 1.000$; $t_{14} = -1.303$, adjusted $P = 0.640$, respectively) or with gradual perturbation ($t_{14} = -1.036$, adjusted $P = 0.952$; gradual 24 h: $t_{14} = -2.281$, adjusted $P = 0.116$, respectively).

Similarly, in the speech feedback conditions, we also assessed whether learning and retention were reliably different than zero. Participants showed learning that was reliably different than zero with abrupt perturbation in the 8-h and 24-h conditions ($t_{12} = -5.072$, adjusted $P = 0.002$; $t_{14} = -8.437$, adjusted $P < 0.001$, respectively), and with gradual perturbation in the 8- and 24-h conditions ($t_{14} = -5.878$, adjusted $P < 0.001$; $t_{14} = -5.765$, adjusted $P < 0.001$, respectively). Likewise, participants showed retention that was reliably different than zero with abrupt perturbation in the 8- and 24-h conditions ($t_{12} = -5.542$, adjusted $P = 0.001$; $t_{14} = -3.566$, adjusted $P = 0.024$, respectively). Retention was also observed with gradual perturbation in both the 8- and 24-h conditions ($t_{14} = -3.709$, adjusted $P = 0.018$; $t_{12} = -6.616$, adjusted $P < 0.001$, respectively). Overall, this analysis showed that while learning was evident both with speech and with noise feedback, retention was evident only with speech feedback, and associated speech errors. When speech was reintroduced during tests of retention, performance switched abruptly to show adaptation, which indicates contextual access to sensorimotor memory.

We also conducted a fine-grained analysis of retention that focused specifically on the very first bin following the reintroduction of speech feedback. Specifically, retention was computed using the average of last two trials of that bin, that is, the very first two trials following the first speech feedback trial. Data beyond 2.5 standard deviations from the mean were excluded from this analysis (<2% of the total trials across conditions). When analyzed in this way, mean retention ± SE was 65 ± 11% and 71 ± 10% of the asymptote for abrupt and gradual conditions, respectively. Values were similar for an analysis of 8- versus 24-h retention (means ± SE: 71 ± 11% and 65 ± 10% of the asymptote respectively). Statistically, there were no differences in retention across experimental conditions (no main effect of abrupt vs. gradual perturbation: $F_{1,54} = 0.157$, $P = 0.693$; no main effect of 8- vs.

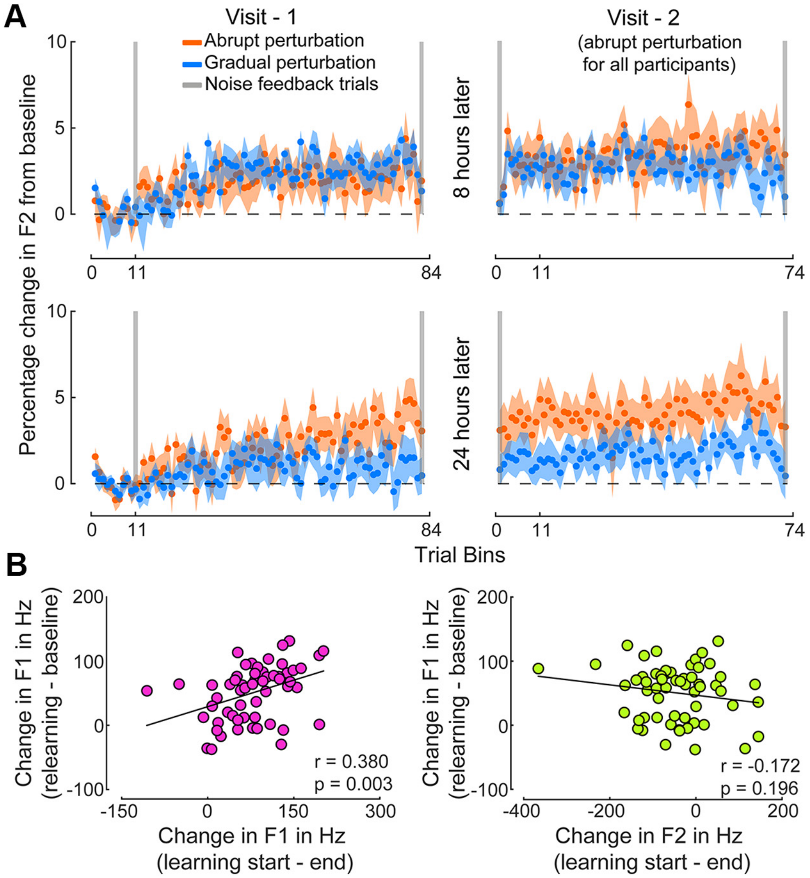

Figure 3. Changes in second formant (F2) do not contribute to first formant (F1) retention. *A*: percentage change in the second formant (F2) from baseline during learning and retention. The horizontal dashed line indicates zero percentage change in F2 from baseline. Each solid circle shows the average value over three consecutive trials (1 bin) and the shaded region gives the standard error. Vertical gray bars indicate bins with speech-modulated noise feedback. During the baseline block, these trials were interspersed with nonperturbed trials, but are combined and shown at the end of baseline block for visualization purposes. *B*: scatter plots showing correlations between retention of F1 and learning of F1 (*bottom left*) or learning of F2 (*bottom right*). Each solid circle indicates data from one participant.

24-h retention interval: $F_{1,54} = 0.195$, $P = 0.660$; no interaction effect of perturbation × retention interval: $F_{1,54} = 2.769$, $P = 0.101$).

Since noise feedback typically results in an increase in speech volume, we conducted an analysis of speech amplitude in noise feedback trials using area under the curve between voice start and end. We assessed amplitude at four time points within the experiment during which noise feedback trials were used: baseline, end of *visit 1*, retention testing, end of *visit 2*. In general, we observed greater amplitude with noise feedback (mean change from baseline ± SE: 12.43 ± 4.88%). However, there were no consistent differences in speech amplitude between conditions. Differences in speech amplitude that could affect the interpretation of the results would relate to differences between end of learning on *visit 1* and retention. We found that while first formant frequency values changed during these noise feedback trials (Fig. 2), speech amplitude values did not ($t_{114} = 1.81$, $P = 0.07$, α corrected for multiple comparisons = 0.008). This argues against the possibility that the formant frequency patterns related to retention arise as a result of changes in speech volume.

Changes in the Second Formant during Learning and Retention

Although the perturbation altered auditory feedback by changing only the first formant, it is possible that learning to compensate for the perturbation could lead to changes in the second formant as well. Consequently, we assessed whether changes in the second formant occurred in association with the applied perturbation. Figure 3*A* shows percentage change in F2 from baseline during both visits, with passage of time (8- and 24-h conditions) and manner of introducing the perturbations (abrupt and gradual conditions). The vertical gray bars indicate noise feedback bins. Although there appear to be differences in F2 for gradual versus abrupt training regimes in the 24 h condition, overall, there were no consistent changes in F2 during the two visits.

We assessed statistically whether changes in F2 associated with learning and retention were reliably different than zero. Participants showed an increase in F2 at the end of learning, but only with a gradual perturbation in the 8-h condition ($t_{14} = 3.214$, adjusted $P = 0.049$). We observed no change in F2 relative to zero at the end of learning a gradual perturbation in the 24-h condition, nor with an abrupt perturbation in the 8- or 24-h conditions (all $|t| < 3.156$, all adjusted $P > 0.056$). We observed no retention in F2, meaning F2 was no different than baseline, with either an abrupt or gradual perturbation or in the 8- or 24-h conditions (all $|t| < 2.07$, all adjusted $P > 0.478$). The changes in F2 during learning relative to baseline were assessed using ANOVA. It was found that there were no systematic differences in F2 in the 8- versus 24-h conditions, nor for abrupt versus gradual introduction of the perturbation (Retention interval: $F_{1,54} = 0.072$, $P = 0.788$; and Perturbation pattern: $F_{1,54} = 0.561$, $P = 0.457$; Retention interval × Perturbation pattern: $F_{1,54} = 3.850$, $P = 0.054$). Similarly, changes in F2 from the start to the end of *visit 2* did not vary with the retention interval nor with the manner of introduction of the perturbation (Retention interval: $F_{1,54} = 0.466$, $P = 0.497$; and Perturbation pattern: $F_{1,54} = 1.289$, $P = 0.261$; Retention interval × Perturbation pattern: $F_{1,54} = 0.494$, $P = 0.485$).

The observation of changes in first formant and marginal changes in second formant frequencies led us to probe the association of these changes and their potential contribution to learning and retention. Figure 3*B* shows scatterplots of the relation between retention as measured in F1, with learning observed in F1 (Fig. 3*B*, *top*) or learning observed in F2 (Fig. 3*B*, *bottom*). It can be seen that only learning related changes in F1, but not in F2, account for the observed retention in F1.

These relationships were assessed quantitatively by computing correlation coefficients. Given that each of our learning and retention measures were normalized by the same baseline values (to homogenize the comparisons), we instead computed learning by subtracting last bin value from the first bin value of the learning block (speech feedback trials), and retention by subtracting first bin value of *visit 2* from the average value of the baseline block (speech feedback trials). We assessed correlations involving changes in F1 and F2 during learning and retention (F1 learning and F1 retention: Pearson $r = 0.380$, $P = 0.003$; F1 learning and F2 learning: Pearson $r = -0.345$, $P = 0.007$; F2 learning and F1 retention: Pearson $r = -0.172$, $P = 0.196$; F2 retention and F1 retention: Pearson $r = -0.02$, $P = 0.875$; F2 learning and F2 retention: Pearson $r = 0.056$, $P = 0.671$; F1 learning and F2 retention: Pearson $r = 0.004$, $P = 0.973$). Among these, only the relationships between F1 learning and F1 retention, and between F1 learning and F2 learning were reliable.

Next, to assess the contribution of changes in the first and second formants to retention, we used a multiple linear regression analysis with one dependent variable (retention in F1), and three predictors (learning in F1, learning in F2, and retention in F2). The model, overall, accounted for a significant amount of the variance in F1 retention ($F_{3,54} = 3.100$, $P = 0.034$). However, only F1 learning was a significant predictor of F1 retention ($P = 0.009$), and not the changes in F2 (during learning: $P = 0.739$, or during retention: $P = 0.874$). Our regression model and correlation analysis thus indicated that there was no contribution of changes in F2 during learning and retention to the observed retention in F1.

Speech Motor Learning Leads to an Acoustical Shift into a Novel Part of the Vowel Space

To understand where in the vowel space the newly learned movements stabilize, we conducted a control study using the vowel /ɛ/ (the same vowel that we tested for in primary studies), with respect to already well-learned vowels (/æ/ and /ɪ/). In this control study, data were recorded for a baseline block involving three stimuli (*dep*, *dap*, and *dip* in pseudorandom order) with no perturbation. This block was followed by a learning block in which the first formant frequency of each repetition of *dep* was gradually increased to 30%. The question was whether the new learning overlaps with neighboring vowel or occupies a novel position within the vowel space.

Figure 4*A* shows baseline and learning block data for the vowel /ɛ/ (*dep*). The percentage change in F1 relative to baseline is shown with no perturbation trials to the left of the vertical dotted line and learning trials to the right. Figure 4*B* shows mean values for each participant, for vowels /ɛ/, /æ/, and /ɪ/ (*dep, dap, dip,* respectively) during baseline and for the vowel /ɛ/ (*dep*) at the end of learning. The ellipses show the 95 percentile values for each vowel cluster. The arrows in Fig. 4, *B* and *C* provide a schematic representation of vector differences between these clusters. Shifts relative to the vowel /ɛ/ under baseline conditions are subscripted BASELINE in the figure and shifts relative to the /ɛ/ at the end are learning are subscripted LEARN. Fig. 4, *D–F*, quantifies the mean direction and magnitude of the individual acoustical changes. It is seen that the vowel /ɛ/ (*dep*) at the end of learning shifted both in direction and magnitude from the same vowel during baseline (Fig. 4*D*). This change in /ɛ/ following learning resulted in a vectorial shift of the vowel /æ/ from /ɛ/ at the end of learning compared with that from /ɛ/ at baseline (Fig. 4*E*). Similarly, the change in /ɛ/ resulted in a vectorial shift of the vowel /ɪ/ from /ɛ/ at the end of learning compared with that from /ɛ/ at baseline (Fig. 4*F*). Overall, it can be seen that at the end of learning, the vowel /ɛ/ shifts to a region of the vowel space which is not overlapping with that of previously well-learned vowels /ɛ/, /æ/, and /ɪ/.

The acoustical differences in this control study were tested statistically. The analysis showed that learning a new speech-acoustical map indeed resulted in an acoustical shift in /ɛ/ compared with its baseline (Fig. 4*D*, vector from /ɛ/ baseline to /ɛ/ learning was different in magnitude: $t_{13} = 11.889$, $P < 0.001$; and direction: $t_{13} = 15.088$, $P < 0.001$). Further evidence of this shift, as assessed by vectors from /ɛ/ to /æ/ at the end of learning versus /ɛ/ to /æ/ at baseline, was observed in terms of a shift in vector magnitude (Fig. 4*E*, $t_{13} = -5.200$, $P < 0.001$) but not direction ($t_{13} = -1.293$, $P = 0.218$). In addition, vectors from /ɛ/ to /ɪ/ (end of learning vs. baseline) showed a shift in magnitude (Fig. 4*F*, $t_{13} = 2.933$, $P = 0.011$) and in direction ($t_{13} = 4.568$, $P < 0.001$).

Directional analyses of the data shown in Fig. 4, *B* and *C* were repeated using circular statistics via the *CircStat* MATLAB toolbox (34). The direction from baseline to end of learning for /ɛ/ was different than zero (mean ± SE: 3.12 ± 0.19 radians). The direction from baseline /ɛ/ to baseline /æ/ (5.77 ± 0.12) was not different than that from the end of learning /ɛ/ to baseline /æ/ (5.92 ± 0.07). However, the direction from baseline /ɛ/ to baseline /ɪ/ (2.34 ± 0.02) was different than that from end of learning /ɛ/ to baseline /ɪ/ (2.03 ± 0.05 radians). There were no differences in outcome between the analysis reported earlier and that conducted using circular statistical methods. Taken together, these analyses are consistent with the idea that speech motor adaptation leads to a consistent shift in speech acoustics into a novel space different than that occupied by already learned neighboring vowels.

Tests with circular statistics were repeated using the adaptation data from the main experimental manipulation. The focus was on the direction of the speech acoustical shift from baseline to end of learning for /ɛ/, as this vowel was tested both in the main experiment and the control study. In an overall comparison, which pooled data over all experimental conditions, it was found that there was no difference in the direction of shift between the two studies ($F_{1,70} = 2.55$, $P = 0.11$). The observed direction of shift was 3.12 ± 0.19 radians, for the control study (mean ± SE), and 2.83 ± 0.05 radians

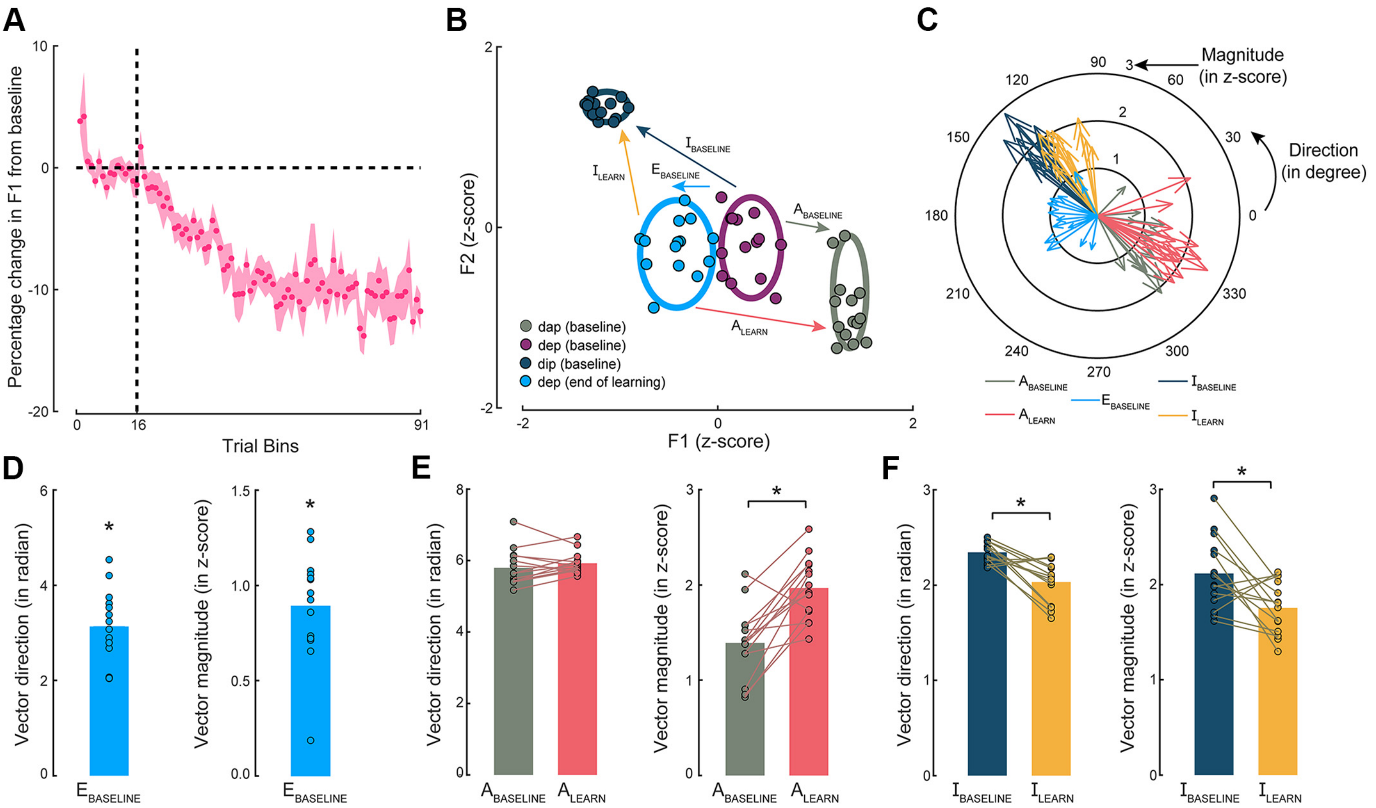

Figure 4. New learning stabilizes within a novel region of the vowel space. *A*: percentage change in first formant frequency during learning. Each bin consists of an average over three consecutive trials. Baseline data (16 bins) until the vertical dashed line, followed by a learning block (75 bins) in which the first formant frequency is shifted up. All utterances here are for the stimulus "dep." The horizontal dashed line indicates 0% change relative to baseline. *B*: map of the vowel space for /Ɛ/ (dep), /æ/ (dap), /ɪ/ (dip) during baseline and for /Ɛ/ (dep) at the end of learning. Each solid circle represents averaged data for 50 trials of that vowel for one participant; each ellipse marks the 95-percentile boundary for a given vowel. Arrows represent changes relative to /Ɛ/ with the convention that the arrow points to the vowel which is represented by the letter. The subscript indicates the experimental condition (e.g., $A_{BASELINE}$ or A_{LEARN}). *C*: polar plot showing vectors from baseline of /Ɛ/ (dep) to each vowel and from /Ɛ/ following learning to /æ/ and /ɪ/. Numbers around the circle indicate direction in degrees and the magnitude is indicated by smaller to larger circles. *D*: comparison of direction and magnitude of vectors from /Ɛ/ (dep baseline) to /Ɛ/ (dep at the end of learning). *E*: comparison of vectorial direction and magnitude from /Ɛ/ (dep baseline) to /æ/ (dap baseline) vs. /Ɛ/ (dep following learning) to /æ/ (dap baseline). *F*: comparison of direction and magnitude of vectors from /Ɛ/ (dep baseline) to /ɪ/ (dip baseline) vs. /Ɛ/ (dep following learning) to /ɪ/ (dip baseline). For plots *D–F*, each solid circle of similar color indicates the value for each participant and $^{*}P$ values less than 0.05.

in the main study. Additional tests were conducted between the control condition and the individual directions observed in each of the gradual versus abrupt perturbation, and 8- versus 24-h retention interval conditions. None of these differences reached statistical significance ($P > 0.05$ in all cases). Thus, the angle of change in F1/F2 space was not found to differ between the main study and the control. This suggests that the findings from the control study, a shift toward a novel part of the vowel space following speech motor learning, likewise apply to the main experimental manipulation.

DISCUSSION

The primary goal of this study was to investigate whether the human speech-auditory system could serve as a model to probe sensorimotor memory retention and retrieval. Specifically, we assessed the durability, and memory access characteristics of newly acquired speech-acoustical memories in the context of speech-motor learning. Separate experimental conditions involved participants learning to compensate for a 30% increase in their first formant (F1) frequency following either an abruptly or gradually introduced perturbation and then returning for a test of retention either 8- or 24-h later. We observed that there was extensive retention of the new learning after both 8- and 24-h (∼70% retention) (Fig. 2). The speech memories were insensitive to the manner of acquisition (gradual vs. abrupt shifts in auditory feedback) (Fig. 2). Furthermore, by introducing speech-modulated noise feedback trials, we observed degraded retention in the presence of noise followed by an abrupt switch to an adapted state when speech error feedback was made available. This indicates that two modes of speech processing exist in parallel, one corresponding to habitual speech and the other to adapted speech. The availability of speech auditory feedback acts as a switch to change from the habitual to the adapted state (Fig. 2). Moreover, the changes that we observe from one trial to the next are an order of magnitude greater than the single-trial learning estimates reported in previous work (33). This extremely rapid transition toward asymptotic values after speech was reintroduced is consistent with the idea that these trials capture retention of learning as opposed to relearning (see Fig. 2*A*, *insets*). In addition, as seen in a control study, the new learning stabilized in a novel part of the sound space that did not overlap with the acoustical placement of neighboring vowels (Fig. 4). In summary, this study shows that just a single session of speech acoustical learning results in extensive retention, which is unaltered following an overnight delay, and reveals the contextual dependence of these memories on speech acoustical feedback.

Retention and Contextual Retrieval of Speech-Acoustical Memories

Speech motor learning through perturbation of various speech parameters has been shown previously. For instance, perturbation of vowel frequency, pitch, vowel duration, and intensity has been shown to elicit corrective learning responses (13, 14, 16–18). However, the durability of the learned responses was unknown. Since washout of the new learning might occur outside of laboratory settings as participants engage in conversations, it is possible that speech-motor learning following a single session could be a transitory phenomenon such that no record of learning may exist later. Findings from the current study argue against this possibility, establishing the durability of the new learning after a single session. The new speech-acoustical memory was durable when assessed 8 h following learning. Retention following 24 h was equivalent to that observed following 8 h, indicating no offline sleep-dependent gains.

The present findings of incomplete learning yet near complete retention in the case of speech stand in contrast to findings in the work on limb movement, in which adaptation is complete, as is retention when tested following a delay (35–37). Incomplete learning in speech adaptation may be attributable to imprecise speech-acoustical targets and the lack of explicit processes for error compensation (38, 39). Speech and limb movement differ in the extent of retention when probed without feedback. In the present study, we observed almost no retention whatsoever when speech feedback was masked with noise, whereas in limb movement, substantial retention is observed even in the absence of visual feedback (40). This suggests that speech memory is heavily dependent upon auditory feedback while memory for upper limb movement is significantly somatic.

A notable aspect of the current findings is the dependence of speech-acoustical memory retrieval on the availability of error information in auditory feedback. The role of feedback was assessed in the present study by introducing noise feedback trials at different points in the experimental sequence. Noise feedback trials have been considered essential for the assessment of the extent of adaptation in the speech-motor learning literature as they provide a measure of feed-forward control, that is, a measure of performance in the absence of online corrections (13, 41, 42). Consistent with these studies, our findings show adaptation by introducing noise feedback trials at the end of learning. However, noise feedback substantially underestimated the extent of retention when introduced at the very beginning of *visit 2* either 8- or 24-h later. Thus, the new speech-acoustical memories were accessible in conjunction with noise feedback only when tested immediately after learning, but not 8-h later. Nevertheless, these memories were present and retrievable one trial after the reintroduction of speech feedback and resulting speech errors at the very start of *visit 2*, indicating the contextual dependence of memory retrieval on the availability of speech error feedback.

The finding that a context cue enables access to the speech acoustical memory is consistent with findings in the limb motor literature (43). In the present study, limited retention was observed in the presence of noise feedback trials, whereas near complete retention of prior learning was evident immediately upon the reintroduction of speech auditory feedback. In work on upper limb movement, in tests of after-effects following force field adaptation, retention was observed when participants held the robotic handle, whereas there was no evidence of retention when participants performed after-effect movements without the robotic handle. Thus, in both cases, evidence of retention is context-specific. In speech, with the passage of time after initial learning, speech sounds return to their habitual state, as observed with noise feedback trials at the beginning of the retention test. However, with the introduction of speech-auditory feedback, it is evident that there is near-complete retention of prior learning. Thus, the adapted learning exists in parallel with the habitual state and is accessible with the appropriate context.

It is possible that the stabilization of the new learning in a novel region within the vowel space could provide the distinctiveness needed to separate the adapted state from the vowel space of habitual speech. This separation could, in turn, provide a reference that, in conjunction with speech feedback, enables a rapid switch back to the adapted state.

Retention of newly acquired sensorimotor memory has been consistently observed in studies involving limb movement (25, 35, 36, 44). The dependence of retention of sensorimotor memory on time elapsed since acquisition and upon sleep has been considered in several limb motor studies (20, 21). In particular, the role of sleep in enhancing consolidation of sensorimotor memory has been proposed to be dependent on the amount of practice and the task under investigation (20, 21). For instance, sleep-dependent gains in motor performance were greater in a visual search task having fewer trials during practice when compared with that with more trials (45). Although sleep is known to enhance retention in a sequential finger-tapping task (46), sleep-dependent consolidation was shown to occur only for a sequential finger-tapping task but not for a visuomotor adaptation task (47). In contrast, a wakeful resting interval of only 10 min following learning was shown to enhance motor performance to a level that was at par with performance observed following uninterrupted and fragmented sleep episodes for a finger-tapping task (47). In the case of force field adaptation, participants deprived of sleep for 24 h following learning of a force field showed retention similar to those who underwent routine sleep following the learning task (48), all of which indicate that the consolidation of procedural memory is highly task dependent. Findings from our study further extend support to the notion that consolidation in the context of new speech-acoustical memories is dependent on passage of time but not necessarily sleep.

Speech Motor Learning May Stabilize in a Novel Region of the Acoustical Space

The control study showed that the speech motor learning paradigm in the present article led participants to produce a new vowel that did not overlap with well-learned vowels in the acoustical neighborhood. This finding is in contrast with that observed by Daliri et al. (49) who reported a response to perturbation along an axis defined by well-learned neighboring vowels rather than directly opposite to the perturbation. Differences in experimental design may account for this. In particular, Daliri et al. (49) focused on assessing the compensatory response by examining formant data toward the end of a prolonged vowel production (∼400 ms), which they show to be influenced by online auditory feedback-driven processes. The use of online feedback presumably underlies the alignment of vowel production with the perceptual targets of the existing neighboring vowels. This could be observed in that report by changes in both F1 and F2 frequencies in the later phase of a prolonged vowel. In contrast, the present study was primarily designed to probe the feedforward component within the early phase of vowel generation to assess speech-motor learning. This process presumably taps into a low-level sensorimotor correction opposite in direction to the applied perturbation. In an experimental design comparable with the one in this study, Lametti et al. (50) showed changes in F1 and F2 frequencies quite similar to our findings. That is, the distance decreased between the newly learned vowel production and neighboring vowels, apparently suggesting the alignment of newly learned vowel with nearby vowels. However, the direction of this change, both in our data and in Lametti et al.'s, is altered such that the new production does not overlap with existing vowels. This supports the idea that following a similar speech-motor learning paradigm, participants are led to produce new vowel sounds.

One methodological issue was the use of pseudowords as opposed to real words in the present study. This was done in part to reduce interference with new learning from routine speech outside of the laboratory. It was also done to restrict possible semantic involvement during learning and to ensure that the participants learned novel speech motor plans rather than altering pre-existing ones in the process. Although it is unclear whether this degree of retention would be observed had we chosen to use real words, the present findings show that almost complete retention of prior learning is possible in the context of audio-motor adaptation.

To conclude, we establish that a single session of speech motor adaptation involving formant perturbation leads to retention of learning 8 h later that is unaltered 24-h following acquisition. At both delays, the retention is almost entirely complete, and no offline gains in retention are observed. Speech acoustical memories are found to be insensitive to the manner of their acquisition (abrupt vs. gradual). In addition, the retrieval of these memories is context-specific and dependent upon the availability of speech error feedback. These findings open avenues for the investigation of retention of speech-motor learning that may be used to probe speech-acoustical memories in intact and neurologically compromised nervous systems.

DATA AVAILABILITY

Data can be made available upon request to the authors.

ACKNOWLEDGMENTS

The authors thank Dr. Matthias Franken for assistance in setting up an earlier version of the speech task, Yiyun Hua for assistance with data collection and curation, and Danilo Lombardo, Drs. Wei-Rong Chen, Ilaine Wang, and Douglas Whalen for helpful discussions on the topic.

GRANTS

The authors acknowledge the support from National Institute on Deafness and Other Communication Disorders (NIDCD) Grants R01DC017439 and R01DC022097 (to D.J.O.).

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the authors.

AUTHOR CONTRIBUTIONS

N.R. and D.J.O. conceived and designed research; N.R. performed experiments; N.R. analyzed data; N.R. and D.J.O. interpreted results of experiments; N.R. prepared figures; N.R. and D.J.O. drafted manuscript; N.R. and D.J.O. edited and revised manuscript; N.R. and D.J.O. approved final version of manuscript.

REFERENCES

1. **Krakauer JW**, **Shadmehr R.** Consolidation of motor memory. *Trends Neurosci* 29: 58–64, 2006. doi:10.1016/j.tins.2005.10.003.
2. **Ostry DJ**, **Gribble PL.** Sensory plasticity in human motor learning. *Trends Neurosci* 39: 114–123, 2016. doi:10.1016/j.tins.2015.12.006.
3. **Henriques DYP**, **Cressman EK.** Visuomotor adaptation and proprioceptive recalibration. *J Mot Behav* 44: 435–444, 2012. doi:10.1080/00222895.2012.659232.
4. **Yadav T**, **Tellez OM**, **Francis JT.** Reward-dependent graded suppression of sensorimotor beta-band local field potentials during an arm reaching task in NHP. In: *2022 44th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*. Glasgow, Scotland, United Kingdom: Institute of Electrical and Electronics Engineers Inc., 2022, p. 3123–3126.
5. **Ferezou I**, **Haiss F**, **Gentet LJ**, **Aronoff R**, **Weber B**, **Petersen CCH.** Spatiotemporal dynamics of cortical sensorimotor integration in behaving mice. *Neuron* 56: 907–923, 2007. doi:10.1016/j.neuron.2007.10.007.
6. **Klinzing JG**, **Niethard N**, **Born J.** Mechanisms of systems memory consolidation during sleep. *Nat Neurosci* 22: 1598–1610, 2019. [Erratum in *Nat Neurosci* 22: 1743–1744, 2019]. doi:10.1038/s41593-019-0467-3.
7. **Rao N**, **Parikh PJ.** Fluctuations in human corticospinal activity prior to grasp. *Front Syst Neurosci* 13: 77, 2019. doi:10.3389/fnsys.2019.00077.
8. **Lametti DR**, **Rochet-Capellan A**, **Neufeld E**, **Shiller DM**, **Ostry DJ.** Plasticity in the human speech motor system drives changes in speech perception. *J Neurosci* 34: 10339–10346, 2014. doi:10.1523/JNEUROSCI.0108-14.2014.
9. **Houde JF**, **Jordan MI.** Sensorimotor adaptation in speech production. *Science* 279: 1213–1216, 1998. doi:10.1126/science.279.5354.1213.
10. **Purcell DW**, **Munhall KG.** Compensation following real-time manipulation of formants in isolated vowels. *J Acoust Soc Am* 119: 2288–2297, 2006. doi:10.1121/1.2173514.
11. **Purcell DW**, **Munhall KG.** Adaptive control of vowel formant frequency: evidence from real-time formant manipulation. *J Acoust Soc Am* 120: 966–977, 2006. doi:10.1121/1.2217714.
12. **Lametti DR**, **Smith HJ**, **Watkins KE**, **Shiller DM.** Robust sensorimotor learning during variable sentence-level speech. *Curr Biol* 28: 3106–3113.e2, 2018. doi:10.1016/j.cub.2018.07.030.
13. **Villacorta VM**, **Perkell JS**, **Guenther FH.** Sensorimotor adaptation to feedback perturbations of vowel acoustics and its relation to perception. *J Acoust Soc Am* 122: 2306–2319, 2007. doi:10.1121/1.2773966.
14. **Mitsuya T**, **MacDonald EN**, **Munhall KG**, **Purcell DW.** Formant compensation for auditory feedback with English vowels. *J Acoust Soc Am* 138: 413–424, 2015. doi:10.1121/1.4923154.
15. **Chao SC**, **Daliri A.** Effects of gradual and sudden introduction of perturbations on adaptive responses to formant-shift and formant-clamp

perturbations. *J Speech Lang Hear Res* 66: 1588–1599, 2023. doi:10.1044/2023_JSLHR-21-00435.

16. **Patel R**, **Niziolek C**, **Reilly K**, **Guenther FH.** Prosodic adaptations to pitch perturbation in running speech. *J Speech Lang Hear Res* 54: 1051–1059, 2011. doi:10.1044/1092-4388(2010/10-0162).
17. **Patel R**, **Reilly KJ**, **Archibald E**, **Cai S**, **Guenther FH.** Responses to intensity-shifted auditory feedback during running speech. *J Speech Lang Hear Res* 58: 1687–1694, 2015. doi:10.1044/2015_JSLHR-S-15-0164.
18. **Karlin R**, **Naber C**, **Parrell B.** Auditory feedback is used for adaptation and compensation in speech timing. *J Speech Lang Hear Res* 64: 3361–3381, 2021. doi:10.1044/2021_JSLHR-21-00021.
19. **Rochet-Capellan A**, **Ostry DJ.** Simultaneous acquisition of multiple auditory-motor transformations in speech. *J Neurosci* 31: 2657–2662, 2011. doi:10.1523/JNEUROSCI.6020-10.2011.
20. **King BR**, **Hoedlmoser K**, **Hirschauer F**, **Dolfen N**, **Albouy G.** Sleeping on the motor engram: the multifaceted nature of sleep-related motor memory consolidation. *Neurosci Biobehav Rev* 80: 1–22, 2017. doi:10.1016/j.neubiorev.2017.04.026.
21. **Schmid D**, **Erlacher D**, **Klostermann A**, **Kredel R**, **Hossner EJ.** Sleep-dependent motor memory consolidation in healthy adults: a meta-analysis. *Neurosci Biobehav Rev* 118: 270–281, 2020. doi:10.1016/j.neubiorev.2020.07.028.
22. **Huang VS**, **Shadmehr R.** Persistence of motor memories reflects statistics of the learning event. *J Neurophysiol* 102: 931–940, 2009. doi:10.1152/jn.00237.2009.
23. **Orban De Xivry JJ**, **Criscimagna-Hemminger SE**, **Shadmehr R.** Contributions of the motor cortex to adaptive control of reaching depend on the perturbation schedule. *Cereb Cortex* 21: 1475–1484, 2011. doi:10.1093/cercor/bhq192.
24. **Orban De Xivry JJ**, **Lefèvre P.** Formation of model-free motor memories during motor adaptation depends on perturbation schedule. *J Neurophysiol* 113: 2733–2741, 2015. doi:10.1152/jn.00673.2014.
25. **Hussain SJ**, **Morton SM.** Perturbation schedule does not alter retention of a locomotor adaptation across days. *J Neurophysiol* 111: 2414–2422, 2014. doi:10.1152/jn.00570.2013.
26. **Coltman SK**, **van Beers RJ**, **Medendorp WP**, **Gribble PL.** Sensitivity to error during visuomotor adaptation is similarly modulated by abrupt, gradual, and random perturbation schedules. *J Neurophysiol* 126: 934–945, 2021. doi:10.1152/jn.00269.2021.
27. **Alhussein L**, **Hosseini EA**, **Nguyen KP**, **Smith MA**, **Joiner WM.** Dissociating effects of error size, training duration, and amount of adaptation on the ability to retain motor memories. *J Neurophysiol* 122: 2027–2042, 2019. doi:10.1152/jn.00387.2018.
28. **Cai S**, **Boucek M**, **Ghosh SS**, **Guenther FH**, **Perkell JS.** A system for online dynamic perturbation of formant trajectories and results from perturbations of the Mandarin Triphthong/iau. In: *8th International Seminar of Speech Production*. Strasbourg, 2008, p. 65–68.
29. **McAuliffe M**, **Socolof M**, **Mihuc S**, **Wagner M**, **Sonderegger M.** Montreal Forced Aligner: trainable text-speech alignment using Kaldi. *Proc Interspeech* 2017: 498–502, 2017. doi:10.21437/Interspeech.2017-1386.
30. **Boersma P**, **Weenink D.** Praat: Doing Phonetics by Computer (Online). Praat, 1992. praat.org.
31. **Boersma P**, **Van Heuven V.** *Speak and unSpeak with PRAAT*. Glot International, 2001, p. 341–347.
32. **Kapsner-Smith MR**, **Abur D**, **Eadie TL**, **Stepp CE.** Test–retest reliability of behavioral assays of feedforward and feedback auditory–motor control of voice and articulation. *J Speech Lang Hear Res* 67: 34–48, 2024. doi:10.1044/2023_JSLHR-23-00038.
33. **Hantzsch L**, **Parrell B**, **Niziolek CA.** A single exposure to altered auditory feedback causes observable sensorimotor adaptation in speech. *eLife* 11: e73694, 2022. doi:10.7554/elife.73694.
34. **Berens P.** CircStat: a MATLAB toolbox for circular statistics. *J Stat Soft* 31: 1–21, 2009. doi:10.18637/jss.v031.i10.
35. **Kumar N**, **Manning TF**, **Ostry DJ.** Somatosensory cortex participates in the consolidation of human motor memory. *PLoS Biol* 17: e3000469, 2019. doi:10.1371/journal.pbio.3000469.
36. **Darainy M**, **Manning TF**, **Ostry DJ.** Disruption of somatosensory cortex impairs motor learning and retention. *J Neurophysiol* 130: 1521–1528, 2023. doi:10.1152/jn.00231.2023.
37. **Hadjiosif AM**, **Morehead JR**, **Smith MA.** A double dissociation between savings and long-term memory in motor learning. *PLoS Biol* 21: e3001799, 2023. doi:10.1371/journal.pbio.3001799.
38. **Kim KS**, **Max L.** Speech auditory-motor adaptation to formant-shifted feedback lacks an explicit component: reduced adaptation in adults who stutter reflects limitations in implicit sensorimotor learning. *Eur J Neurosci* 53: 3093–3108, 2021. doi:10.1111/ejn.15175.
39. **Munhall KG**, **MacDonald EN**, **Byrne SK**, **Johnsrude I.** Talkers alter vowel production in response to real-time formant perturbation even when instructed not to compensate. *J Acoust Soc Am* 125: 384–390, 2009. doi:10.1121/1.3035829.
40. **Ebrahimi S**, **Ostry DJ.** The human somatosensory cortex contributes to the encoding of newly learned movements. *Proc Natl Acad Sci USA* 121: e2316294121, 2024. doi:10.1073/pnas.2316294121.
41. **MacDonald EN**, **Raufer S.** Speech production in amplitude-modulated noise. *J Acoust Soc Am* 133: 3519–3519, 2013. doi:10.1121/1.4806308.
42. **Maas E**, **Mailend M-L**, **Guenther FH.** Feedforward and feedback control in apraxia of speech: effects of noise masking on vowel production. *J Speech Lang Hear Res* 58: 185–200, 2015. doi:10.1044/2014_JSLHR-S-13-0300.
43. **Cothros N**, **Wong JD**, **Gribble PL.** Are there distinct neural representations of object and limb dynamics? *Exp Brain Res* 173: 689–697, 2006. doi:10.1007/s00221-006-0411-0.
44. **Rao N**, **Mehta N**, **Patel P**, **Parikh PJ.** Effects of aging on conditional visuomotor learning for grasping and lifting eccentrically weighted objects. *J Appl Physiol (1985)* 131: 937–948, 2021. doi:10.1152/japplphysiol.00932.2020.
45. **Censor N**, **Karni A**, **Sagi D.** A link between perceptual learning, adaptation and sleep. *Vision Res* 46: 4071–4074, 2006. doi:10.1016/j.visres.2006.07.022.
46. **Lutz ND**, **Wolf I**, **Hübner S**, **Born J**, **Rauss K.** Sleep strengthens predictive sequence coding. *J Neurosci* 38: 8989–9000, 2018. doi:10.1523/JNEUROSCI.1352-18.2018.
47. **Doyon J**, **Korman M**, **Morin A**, **Dostie V**, **Tahar AH**, **Benali H**, **Karni A**, **Ungerleider LG**, **Carrier J.** Contribution of night and day sleep vs. simple passage of time to the consolidation of motor sequence and visuomotor adaptation learning. *Exp Brain Res* 195: 15–26, 2009. doi:10.1007/s00221-009-1748-y.
48. **Donchin O**, **Sawaki L**, **Madupu G**, **Cohen LG**, **Shadmehr R.** Mechanisms influencing acquisition and recall of motor memories. *J Neurophysiol* 88: 2114–2123, 2002. doi:10.1152/jn.2002.88.4.2114.
49. **Daliri A**, **Chao SC**, **Fitzgerald LC.** Compensatory responses to formant perturbations proportionally decrease as perturbations increase. *J Speech Lang Hear Res* 63: 3392–3407, 2020. doi:10.1044/2020_JSLHR-19-00422.
50. **Lametti DR**, **Smith HJ**, **Freidin PF**, **Watkins KE.** Cortico-cerebellar networks drive sensorimotor learning in speech. *J Cogn Neurosci* 30: 540–551, 2018. doi:10.1162/jocn_a_01216.