Report

Current Biology

Tracing the Trajectory of Sensory Plasticity across Different Stages of Speech Learning in Adulthood

Highlights

- Adults were trained to perceptually master a difficult non-native phonetic contrast
- Sensory and perceptual plasticity emerge at different timescales
- Training modifies sensory encoding at a slower timescale relative to perception
- Perceptual and sensory gains are retained beyond the cessation of training

Authors

Rachel Reetzke, Zilong Xie, Fernando Llanos, Bharath Chandrasekaran

Correspondence

bchandra@utexas.edu

In Brief

Reetzke et al. show that as adults are trained to categorize non-native speech sounds, sensory encoding of non-native speech sound patterns improves only after an expert level of behavioral performance. Training-induced changes in sensory encoding relate to behavioral performance and endure beyond the period of training.

Reetzke et al., 2018, Current Biology *28*, 1419–1427
May 7, 2018
https://doi.org/10.1016/j.cub.2018.03.026

CellPress

Current Biology
Report

Tracing the Trajectory of Sensory Plasticity across Different Stages of Speech Learning in Adulthood

Rachel Reetzke,[1] Zilong Xie,[1] Fernando Llanos,[1] and Bharath Chandrasekaran[1,2,3,4,5,6,*]
[1]Department of Communication Sciences and Disorders, The University of Texas at Austin, Austin, TX 78712, USA
[2]Department of Psychology, The University of Texas at Austin, Austin, TX 78712, USA
[3]Department of Linguistics, The University of Texas at Austin, Austin, TX 78712, USA
[4]Institute for Neuroscience, The University of Texas at Austin, Austin, TX 78712, USA
[5]Institute for Mental Health Research, The University of Texas at Austin, Austin, TX 78712, USA
[6]Lead Contact
*Correspondence: bchandra@utexas.edu
https://doi.org/10.1016/j.cub.2018.03.026

SUMMARY

Although challenging, adults can learn non-native phonetic contrasts with extensive training [1, 2], indicative of perceptual learning beyond an early sensitivity period [3, 4]. Training can alter low-level sensory encoding of newly acquired speech sound patterns [5]; however, the time-course, behavioral relevance, and long-term retention of such sensory plasticity is unclear. Some theories argue that sensory plasticity underlying signal enhancement is immediate and critical to perceptual learning [6, 7]. Others, like the reverse hierarchy theory (RHT), posit a slower time-course for sensory plasticity [8]. RHT proposes that higher-level categorical representations guide immediate, novice learning, while lower-level sensory changes do not emerge until expert stages of learning [9]. We trained 20 English-speaking adults to categorize a non-native phonetic contrast (Mandarin lexical tones) using a criterion-dependent sound-to-category training paradigm. Sensory and perceptual indices were assayed across operationally defined learning phases (novice, experienced, over-trained, and 8-week retention) by measuring the frequency-following response, a neurophonic potential that reflects fidelity of sensory encoding, and the perceptual identification of a tone continuum. Our results demonstrate that while robust changes in sensory encoding and perceptual identification of Mandarin tones emerged with training and were retained, such changes followed different timescales. Sensory changes were evidenced and related to behavioral performance only when participants were over-trained. In contrast, changes in perceptual identification reflecting improvement in categorical percept emerged relatively earlier. Individual differences in perceptual identification, and not sensory encoding, related to faster learning. Our findings support the RHT—sensory plasticity accompanies, rather than drives, expert levels of non-native speech learning.

RESULTS AND DISCUSSION

Sound-to-Category Training Paradigm

We trained 20 native English-speaking adults using a criterion-dependent sound-to-category training paradigm (Figure 1; STAR Methods) [13]. As shown in Figures 2A and 2B, each participant was monitored across three operationally defined learning phases until criterion behavioral performance was achieved and maintained. Participants were considered novice at the first training session. Participants took 4–13 days (M = 7.10, SD = 2.81) to reach the experienced learning phase, defined as maintaining behavioral accuracy comparable to native Chinese participants (>90% accuracy) for 3 consecutive days. Participants were then over-trained for 10 additional days beyond the experienced phase to ensure stability in behavioral categorization. 8 weeks post-training, we evaluated retention of behavioral performance [14]. To further test participants' behavioral mastery of Mandarin lexical tone categorization, at each phase we probed learning through two secondary tasks (Figure 1B): speech categorization under a dual-task constraint and generalization to untrained stimuli.

To examine training-induced changes in perceptual identification, we assayed identification accuracy and reaction times for tones drawn randomly from a seven-step tone continuum that ranged from the level to the rising Mandarin lexical tone (Figure 1C). To examine training-induced changes in sensory encoding of non-native speech sound patterns, we measured neural tracking of the four Mandarin lexical tone fundamental frequency (F0) contours using the frequency-following response (FFR) (Figure 1D), a pre-attentive measure of synchronous sound-evoked neural activity that encodes acoustic details of the incoming stimulus along the early auditory pathway (see STAR Methods) [15–17]. English learners' performance on all tasks was compared to native Chinese participants whose categorization performance was used to establish the training criterion.

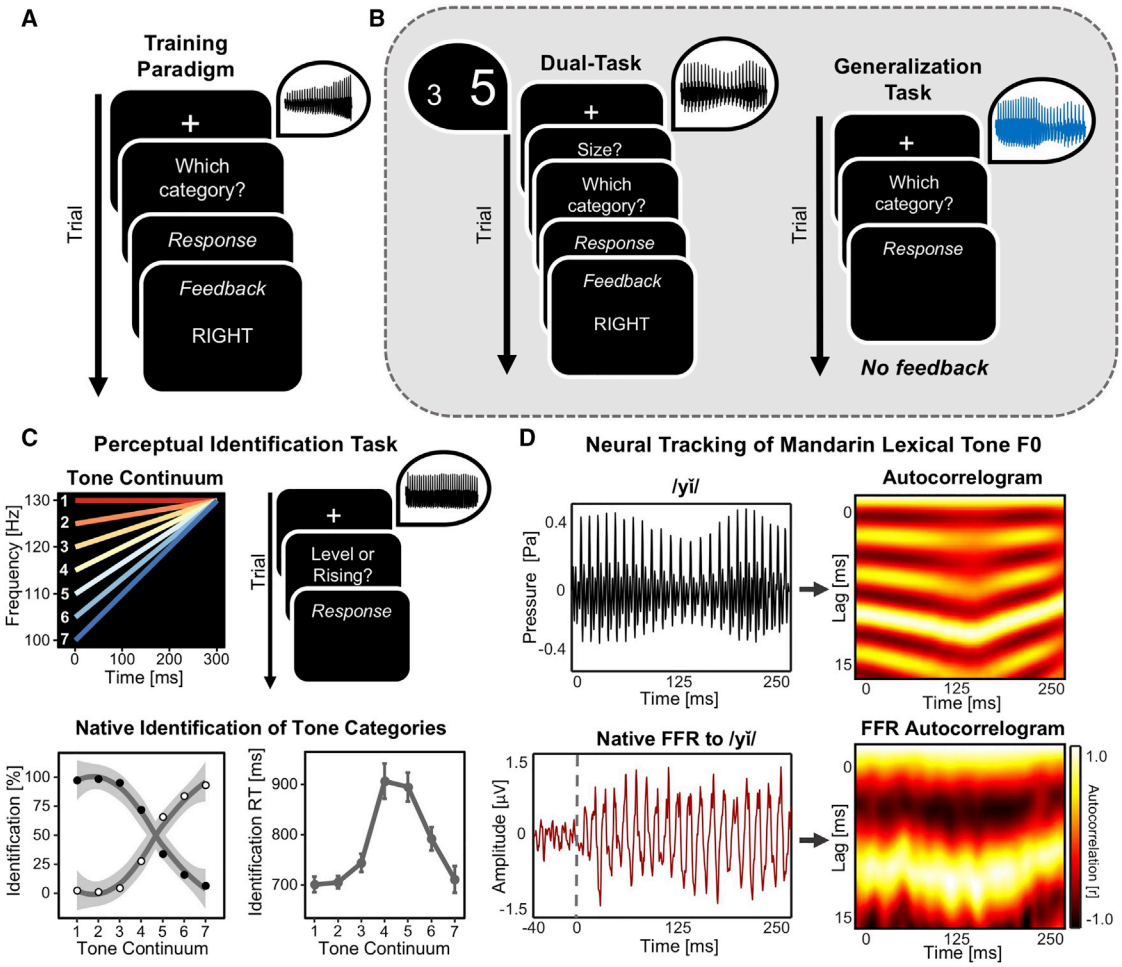

Figure 1. Experimental Methods

(A) Trial procedure for the sound-to-category training task. Each trial began with a fixation cross in the center of the screen for 750 ms. A Mandarin lexical tone was presented for a fixed duration (440 ms). Participants were given unlimited time to categorize the tone into categories 1, 2, 3, or 4. Corrective feedback (1,000 ms) was presented 500 ms following participants' response.

(B) Left: Training blocks involving the dual-task design required participants to make "value" and "size" judgment responses while categorizing the Mandarin lexical tones. Right: In a separate generalization task, participants were instructed to categorize stimuli produced by untrained speakers (denoted by the blue waveform).

(C) Perceptual identification task. Top: The seven-step tone continuum from the level (red) to the rising (blue) Mandarin lexical tone and the trial procedure used to probe perceptual identification of tone categories (see STAR Methods). Bottom: The average identification function from native Chinese participants (n = 13) and identification reaction times for tone identification. Closed and open circles correspond to mean identification percentage for the level and rising Mandarin lexical tones, respectively. Despite the continuous acoustic change, native Chinese participants exhibit a steep perceptual identification slope near the category boundary at the midpoint of the continuum (tone token 4) and are slower to label stimuli near the boundary [10–12]. Shaded areas and error bars denote ± SEM.

(D) Neural tracking of Mandarin lexical tone F0. Top: Waveform and autocorrelogram of an example Mandarin lexical tone. Bottom: Corresponding frequency-following response (FFR) and autocorrelogram from a native Chinese participant. The autocorrelogram provides visualization of autocorrelation over a 40-ms sliding window and allows an estimation of the extent to which the FFR follows F0 changes characterizing the Mandarin tone stimulus (see STAR Methods). The colors represent the strength range of the correlation from high (white; value = 1) to low (dark red; value = −1).

Speech Categorization under Dual-Task Constraint

A standard for determining mastery of perceptual learning is to examine the extent to which that behavior is automatic, or maintained while performing another task in parallel under a dual-task constraint [18–20]. As demonstrated in Figure 2C, initially learners' speech categorization under dual-task constraint significantly differed from native participants (accuracy: $b = -0.399$, SE = 0.036, $t = -11.117$, $p < 0.001$; reaction time: $b = 586.100$, SE = 192.840, $t = 3.039$, $p = 0.004$). However, once learners reached the experienced phase, speech categorization was not statistically different from native participants (accuracy: $b = -0.037$, SE = 0.036, $t = -1.028$, $p = 0.307$; reaction time: $b = 260.200$, SE = 192.840, $t = 1.349$, $p = 0.183$). This high level of performance was maintained at the over-trained phase (accuracy: $b = -0.018$, SE = 0.036, $t = -0.490$, $p = 0.626$; reaction time: $b = 61.000$, SE = 192.840, $t = 0.316$, $p = 0.753$) and retained 8 weeks post-training (accuracy: $b = -0.038$, SE = 0.036, $t = -1.046$, $p = 0.299$; reaction time: $b = 67.150$, SE = 192.840, $t = 0.348$, $p = 0.729$).

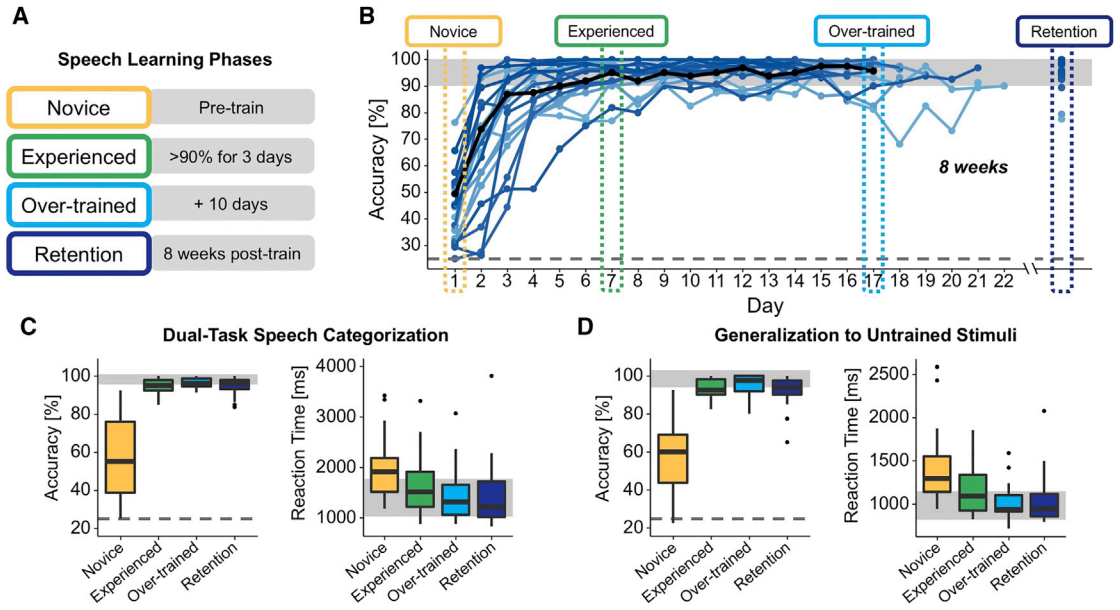

Figure 2. Operationally Defined Speech-Learning Phases and Behavioral Results

(A) The operational definitions for the different speech-learning phases.

(B) Subject-by-day learning curves (n = 20) from the speech-training task. The gray rectangle denotes ± SEM for Chinese participants (target criterion for learners); the dashed gray line indicates chance level (25% accuracy) for Mandarin lexical tone categorization. The emphasized black learning curve is a representative participant tracked across all learning phases.

(C and D) Accuracy and median reaction time results across learning phases for (C) speech categorization under dual-task constraint and (D) generalization to untrained stimuli. The center line on each boxplot denotes the median accuracy or reaction time, the edges of the box indicate the 25th and 75th percentiles, and the whiskers extend to data points that lie within 1.5× the interquartile range. Points outside this range represent outliers. Gray rectangles denote ± SEM for Chinese participants.

Generalization to Untrained Stimuli

A critical test of the mastery of perceptual learning is the extent to which learning generalizes beyond trained stimuli [14]. As shown in Figure 2D, novice learners significantly differed in categorization of untrained stimuli relative to native participants (accuracy: b = −0.400, SE = 0.036, t = −11.124, p < 0.001; reaction time: b = 444.160, SE = 99.750, t = 4.453, p < 0.001). Once learners reached the experienced phase, performance was not statistically different from native participants (accuracy: b = −0.056, SE = 0.036, t = −1.565, p = 0.121; reaction time: b = 170.720, SE = 99.750, t = 1.711, p = 0.092). Maintenance of performance was observed at the over-trained learning phase (accuracy: b = −0.039, SE = 0.036, t = −1.078, p = 0.287; reaction time: b = 35.770, SE = 99.750, t = 0.359, p = 0.721), and retained 8 weeks post-training (accuracy: b = −0.068, SE = 0.036, t = −1.878, p = 0.064; reaction time: b = 59.850, SE = 99.750, t = 0.600, p = 0.551).

Training-Induced Changes in Perceptual Identification of Tone Categories

At each learning phase, we measured the extent to which perceptual identification became more categorical as a function of learning phase by measuring the slope of the tone-identification labeling curve (perceptual slope) and the peak of identification reaction times (peak RT; see STAR Methods). Tone-identification labeling curves and reaction times are shown for each learning phase relative to native performance in Figures 3A and 3D, respectively. Before training, novice learners exhibited a shallower perceptual slope and invariant identification reaction times across the tone continuum, compared to native participants (perceptual slope: b = −1.226, SE = 0.306, t = −4.003, p < 0.001; peak RT: b = −150.140, SE = 71.790, t = −2.091, p = 0.039). However, as shown in Figures 3B and 3E, once learners reached the experienced phase, they demonstrated a steeper perceptual slope and slowing of reaction times at the categorical boundary; this performance was not statistically different from native participants (perceptual slope: b = −0.278, SE = 0.306, t = −0.907, p = 0.367; peak RT: b = −79.810, SE = 71.790, t = −1.112, p = 0.270). Perceptual identification of tone categories was maintained at the over-trained learning phase (perceptual slope: b = −0.112, SE = 0.306, t = −0.367, p = 0.715; peak RT: b = −31.590, SE = 71.790, t = −0.440, p = 0.661), and retained after 8 weeks post-training (perceptual slope: b = −0.419, SE = 0.306, t = −1.369, p = 0.175; peak RT: b = 58.680, SE = 71.790, t = 0.817, p = 0.416).

Training-Induced Changes in Sensory Encoding of Mandarin Lexical Tone F0 Contours

At each learning phase, we measured the FFR to assess changes in neural tracking of the F0 of each of the four Mandarin lexical tones using two well-established metrics: peak autocorrelation, which reflects the robustness of neural phase locking to the F0 contour, and stimulus-to-response correlation, which reflects neural fidelity of F0 tracking [15] (see STAR Methods). Analyses focused on neural tracking of Mandarin tone F0 contours across learning phases as this is

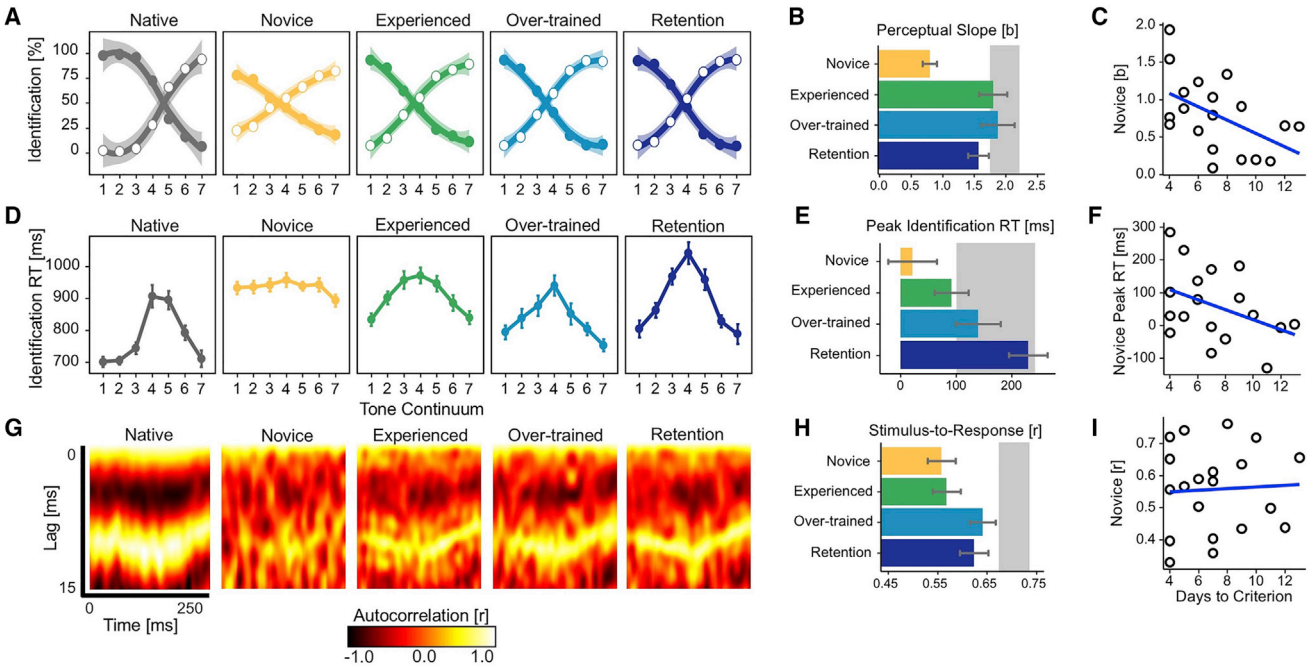

Figure 3. Training-Induced Changes in Perceptual Identification of Tone Categories and Sensory Encoding of Mandarin Lexical Tone F0 Contours

(A) Perceptual identification functions across all learning phases relative to native Chinese participants. Closed and open circles correspond to mean identification percentage for the level and rising Mandarin lexical tones, respectively. Shading denotes ± SEM.

(B) Comparison of steepness of the perceptual identification slope across learning phases. The gray rectangle denotes ± SEM for native Chinese performance. Error bars denote ± SEM.

(C) Individuals with a sharper perceptual slope at the novice learning phase require fewer days of training to reach the criterion level ($r_s = -0.519$, $p = 0.023$).

(D) Identification reaction times for tone identification. Error bars denote ± SEM. See also Table S1.

(E) Comparison of peak of identification reaction time (or slowing) at the categorical boundary. The gray rectangle denotes ± SEM for native Chinese performance. Error bars denote ± SEM.

(F) Relationship between novice peak of identification reaction time and days to reach training criterion ($r_s = -0.351$, $p = 0.153$).

(G) Autocorrelogram function to Mandarin lexical tone 3 for a representative native Chinese participant relative to a representative English learner across all four learning phases. Robust improvement in neural phase-locking (peak autocorrelation) to Mandarin tone F0 is not observed until after behavior is stable (at the over-trained learning phase).

(H) Mean neural tracking accuracy of the F0 contour of all Mandarin tone stimuli as reflected by stimulus-to-response correlation across learning phases. Changes in stimulus-to-response correlation are not observed until the over-trained learning phase. For both neural metrics, plasticity is not tone-specific and is retained after 8 weeks of no training. The gray rectangle denotes ± SEM for native Chinese performance. Error bars denote ± SEM.

(I) In contrast to perceptual identification slope, novice stimulus-to-response correlation (as well as peak autocorrelation, not shown) does not significantly predict days to criterion ($r_s = 0.091$, $p = 0.702$).

the dominant cue for tonal recognition in native speakers of Mandarin Chinese [5, 21–24].

A repeated-measures ANOVA on both metrics showed a main effect of learning phase (peak autocorrelation: $F_{3,\ 57} = 3.49$, $p = 0.033$, $\eta_p^2 = 0.16$; stimulus-to-response correlation: $F_{3,\ 57} = 4.28$, $p = 0.013$, $\eta_p^2 = 0.18$) and tone stimulus (peak autocorrelation: $F_{3,\ 57} = 32.99$, $p < 0.001$, $\eta_p^2 = 0.63$; stimulus-to-response correlation: $F_{3,\ 57} = 38.85$, $p < 0.001$, $\eta_p^2 = 0.67$). The interaction between learning phase and tone stimulus did not reach significance for either metric (peak autocorrelation: $F_{9,\ 171} = 1.46$, $p = 0.203$, $\eta_p^2 = 0.07$; stimulus-to-response correlation: $F_{9,\ 171} = 0.77$, $p = 0.589$, $\eta_p^2 = 0.04$).

Planned comparisons showed that gain in peak autocorrelation did not emerge until the over-trained learning phase (Figure 3G) (over-trained versus novice: $t_{19} = 2.05$, $p = 0.054$, $d = 0.20$; experienced versus novice: $t_{19} = 0.42$, $p = 0.679$, $d = 0.04$). Likewise, stimulus-to-response correlation also did not increase until the over-trained learning phase (Figure 3H) (over-trained versus novice: $t_{19} = 3.88$, $p = 0.001$, $d = 0.34$; experienced versus novice: $t_{19} = 0.39$, $p = 0.669$, $d = 0.04$). Gains in neural phase-locking and neural tracking persisted after 8 weeks post-training, as evidenced by stability between the over-trained and retention learning phases (peak autocorrelation: $t_{19} = 0.80$, $p = 0.435$, $d = 0.08$; stimulus-to-response correlation: $t_{19} = 0.59$, $p = 0.562$, $d = 0.07$).

We conducted Welch two-sample t tests (Bonferroni-corrected) to compare learners' sensory encoding of Mandarin lexical tones at each learning phase to native participants. For the peak autocorrelation metric, learners initially differed from native participants in robustness of neural phase locking to the F0 contour of the Mandarin tone stimuli (novice versus native: $t_{30.67} = 2.72$, $p = 0.010$, $d = 0.70$; experienced versus native: $t_{30.30} = 2.60$, $p = 0.014$, $d = 0.65$); however, once learners reached the over-trained phase, neural phase locking was not statistically different

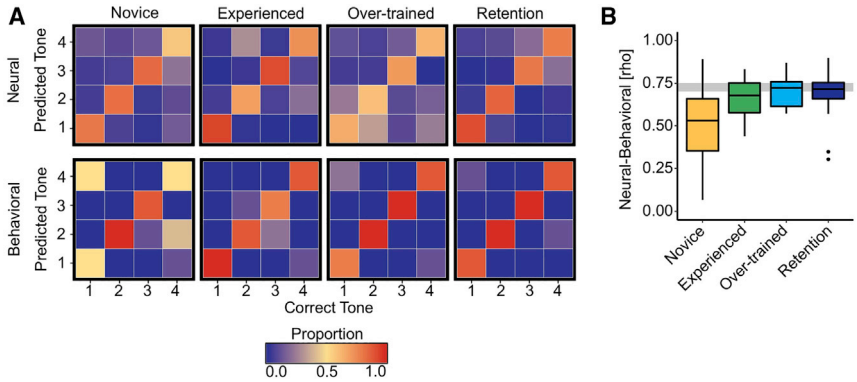

Figure 4. Patterns of Sensory Encoding Predict Patterns of Behavioral Speech Categorization

(A) Neural confusion matrices derived from a Hidden Markov model classifier, which was trained and tested on FFRs to the Mandarin lexical tones (top) and behavioral confusion matrices derived from generalization Mandarin lexical tone categorization performance from a representative English learner. Each matrix corresponds to a learning phase, each row corresponds to predicted Mandarin tone category responses, and each column corresponds to the correct Mandarin tone category. The color of a given cell denotes the proportion of the category-response combination within a given learning phase and ranges from high (red; value = 1.0) to low (blue; value = 0.0). Higher values within the diagonal cells, extending from bottom left to top right corner of the matrix, correspond to correct responses; other cells denote errors.

(B) The relationship between individual neural and behavioral confusions matrices as reflected by Spearman's rho. The gray rectangle denotes ± SEM for native Chinese participants. The center line on each boxplot denotes the median accuracy or reaction time, the edges of the box indicate the 25th and 75th percentiles, and the whiskers extend to data points that lie within 1.5× the interquartile range. Points outside this range denote outliers.

from native participants (over-trained versus native: $t_{30.90} = 1.87$, $p = 0.071$, $d = 0.49$). This high level of performance was retained after 8 weeks post-training (retention versus native: $t_{29.43} = 1.69$, $p = 0.101$, $d = 0.41$). For the stimulus-to-response correlation metric, learners initially differed from native participants in neural tracking of the Mandarin F0 patterns (novice versus native: $t_{28.46} = 3.72$, $p = 0.001$, $d = 0.70$; experienced versus native: $t_{29.22} = 3.40$, $p = 0.002$, $d = 0.65$); however, once learners reached the over-trained phase, stimulus-to-response correlation was not statistically different from native participants (over-trained versus native: $t_{30.54} = 1.66$, $p = 0.108$, $d = 0.34$). Statistical difference between groups was observed at retention (retention versus native: $t_{30.06} = 2.09$, $p = 0.045$, $d = 0.41$).

Sensory and Perceptual Predictors of Days to Speech-Learning Criterion

We investigated the extent to which novice measures of perceptual identification and neural tracking of Mandarin lexical tone F0 patterns predicted the number of training days needed for participants to reach the experienced learning phase. Spearman's rank correlation coefficients revealed that participants' novice perceptual slope negatively related to days to experienced criterion ($r_s = -0.519$, $p = 0.023$), suggesting that individuals with a sharper pre-training perceptual slope required fewer days to criterion (Figure 3C). While peak identification reaction time was not significantly predictive of days to criterion ($r_s = -0.351$, $p = 0.153$), the trend of the correlation was in a similar direction (Figure 3F). Neither of the neural tracking metrics significantly related to days to reach criterion (stimulus-to-response correlation: $r_s = 0.091$, $p = 0.702$; peak autocorrelation: $r_s = -0.111$, $p = 0.641$) (Figure 3I).

Patterns of Sensory Encoding Predict Patterns of Behavioral Speech Categorization

To investigate the relationship between sensory encoding and behavioral categorization of the Mandarin lexical tones across learning phases, we used a Hidden Markov model classifier to decode Mandarin lexical tone categories from the FFRs of each participant (see STAR Methods). Spearman's rank correlation coefficient was calculated to relate the neural confusion matrix provided by the classifier with behavioral patterns of speech categorization from each participant's generalization task performance (Figure 4A). We found that the neural-behavioral relationship between individual patterns of neural decoding and behavioral patterns of untrained Mandarin lexical tone categorization increased across learning phases (Figure 4B) ($F_{3,\ 57} = 6.60$, $p = 0.001$, $\eta_p^2 = 0.26$). The neural-behavioral relationship did not show a statistically relevant improvement until the over-trained learning phase (over-trained versus novice: $t_{19} = 2.79$, $p = 0.012$, $d = 0.90$; experienced versus novice: $t_{19} = 2.06$, $p = 0.054$, $d = 0.59$). Gains in neural-behavioral correspondence persisted after 8 weeks post-training, as evidenced by stability between the over-trained and retention learning phases (over-trained versus retention: $t_{19} = -0.69$, $p = 0.497$, $d = 0.20$).

Welch's two-sample t tests (Bonferroni-corrected) revealed that novice learners exhibited significantly less neural-behavioral correspondence relative to native Chinese participants ($t_{23.84} = -4.05$, $p < 0.001$, $d = 1.31$). Once learners reached the over-trained phase, patterns of sensory encoding related to patterns of behavioral categorization and were not statistically different from native participants (over-trained versus native: $t_{27.24} = -1.35$, $p = 0.188$, $d = 0.44$; experienced versus native: $t_{26.42} = -2.108$, $p = 0.045$, $d = 0.70$). This high level of neural-behavioral correspondence was retained 8 weeks post-training (over-trained versus retention: $t_{30.56} = -1.050$, $p = 0.302$, $d = 0.35$).

General Discussion

We examined the time-course, behavioral relevance, and long-term retention of changes in sensory encoding of non-native speech sound patterns as adults learned to categorize a non-native phonetic contrast across different phases of speech learning. Our results show that learners reached criterion based on native performance, maintained behavioral performance under dual-task constraint, and generalized performance to untrained stimuli, satisfying rigorous standards for perceptual learning mastery [14, 18–20]. While changes in the perceptual identification of tone categories were evidenced at the experienced learning phase, significant changes in sensory encoding of untrained

Mandarin lexical tones emerged only after the over-trained learning phase (Figure 3). Correspondence between patterns of sensory encoding and behavioral categorization of Mandarin lexical tones were also strongly related at this phase. Despite the different timescales for perceptual and sensory plasticity, after 8 weeks post-training, both perceptual changes and sensory-encoding gains were retained. Our findings suggest that sensory enhancement of incoming stimulus features is not critical for early stages of speech perceptual learning. Rather, in line with the RHT, we posit that enhanced sensory encoding observed at a later learning phase is an outcome of perceptual mastery.

Our results diverge from theories that suggest that perceptual learning is primarily driven by sensory processing enhancement [6, 7, 25] but closely align with the RHT, which indicates low-level sensory enhancement emerges at expert stages of perceptual learning [9, 26, 27]. Consistent with RHT, we posit that novice performance is guided by abstract categorical representations [8, 9], likely the result of receptor-field plasticity at higher levels of the sensory processing hierarchy [28]. The emergence of categorical percept may facilitate the top-down guided tuning of lower levels of the auditory hierarchy with the goal of signal enhancement. In line with this idea, animal models have revealed top-down gated sensory enhancement of selective features of incoming acoustic stimuli following different forms of auditory training [29–31]. Our findings suggest that native listeners and over-trained learners may be able to operate at the level of category-based perception and reach down to lower levels of the sensory hierarchy for tuning of behaviorally relevant signals, depending on the nature of the task [32, 33]. Learners who demonstrated better perceptual identification of tone categories (thought to reflect higher-level attention-driven processes [34]) at the novice learning phase took fewer days to reach the learning criterion (Figure 3C). In contrast, measures of neural tracking of non-native speech sound patterns did not relate to faster learning (Figure 3I). Taken together, these studies and our findings suggest that slow changes observed in the sensory encoding of relevant stimulus features (i.e., Mandarin lexical tone F0 contours) may be an outcome of sensory tuning guided by expertise in higher-level categorical perception.

Previous studies using the FFR as a metric have revealed training-induced enhancement in sensory encoding of non-native speech sound patterns in human adults [5, 21, 35]; however, the neurophysiological changes observed have been restricted to specific tones [5] or selective portions of the incoming stimulus [36]. A limitation of previous studies is the large individual differences in learning, likely arising from circumscribed training regimens [5]. In contrast, we show that training-induced improvement in sensory encoding following an individually focused, criterion-driven approach is not tone specific, with overall sensory enhancement observed and retained. This approach allowed for a systematic examination of the neurophysiological changes underlying different phases of perceptual-learning expertise, as all participants were trained to similar levels of behavioral performance. Our paradigm combined high-talker variability stimuli and reinforcement-driven learning via trial-by-trial feedback. These training components have been previously found to direct participant attention to category-relevant acoustic cues and lead to long-lasting behavioral retention [37–39]. Animal models have demonstrated that receptive field properties of neurons in the primary auditory cortex (A1) [40, 41] and the inferior colliculus [42] undergo task-related changes in stimulus-response strength as a result of associative learning. These changes have been connected to dopaminergic projections to A1, activated by associations between incoming stimuli and reinforcers (e.g., reward or punishment) [43, 44]. In line with these findings, a neuroimaging study in adult humans showed that reward-based neural circuitry (i.e., caudate, putamen, and ventral striatum) is activated more in successful learners of a similar speech sound-to-category training task [13]. We posit that the combination of high-talker variability and reinforcement-driven learning, paired with a criterion-driven approach, facilitated behavioral mastery and retention of non-native speech sound learning. This training regimen further led to non-tone-specific improvement in sensory encoding of non-native speech sound patterns.

Despite the different timescales for behavioral and neural plasticity, our findings reveal long-term retention beyond the period of training for both behavioral gains in non-native speech sound categorization and refined sensory encoding of incoming stimulus features. These results are consistent with prior work demonstrating long-term retention of training-induced receptive field reorganization within A1 following extensive auditory discrimination training [43]. Consistent with our findings, sensory encoding gains were observed 8 weeks post-training [41, 45]. In contrast, animal studies investigating auditory and motor training have demonstrated that while cortical plasticity emerges rapidly during learning, such changes may renormalize as behavior stabilizes [46, 47]. This body of work suggests that maintenance and retention of trained performance may be localized to specific neural circuitry rather than in retention of large-scale expansion of tissue in a given neural region [17, 48–50]. While our study cannot speak to the extent to which physiological changes in cortical plasticity were induced or retained, our results suggest that maintenance and retention of sensory encoding of incoming stimulus features, as reflected by the FFR, may be an outcome of training-induced rewired neural circuitry [51, 52]. Furthermore, we observed a strong relationship between patterns of sensory encoding and behavioral categorization at later learning phases and at retention (Figure 4). This strong correspondence suggests that while signal enhancement may not be critical during early stages of perceptual learning, such sensory plasticity is not epiphenomenal. Rather, in line with RHT, we posit that sensory plasticity may be a critical component of behavioral stability and flexibility. For example, under challenging listening conditions, experts may be able to leverage the enhanced signal-to-noise ratio to maintain stability in behavioral performance [53].

Language-specific changes in perception and sensory encoding of speech sounds occur early in life; however, with intensive training adults can learn and retain a difficult non-native phonetic contrasts. Consistent with the RHT, we show that as adults learn non-native phonetic contrasts, sensory encoding is fine-tuned, and accompanies rather than drives expert levels of behavioral perceptual identification. We further provide evidence for training-induced neurophysiological changes in sensory encoding that relate to behavioral stability and endure beyond a period of intensive training. Our findings are in support of an emerging view that auditory perceptual learning is mediated by top-down processes that shape sensory signals through later learning phases [33].

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Sound-to-Category Training Task
  - Dual-Task: Speech Categorization + Numerical Stroop
  - Generalization Task
  - Perceptual Identification Task
  - Electrophysiology
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Evaluation of Changes in Perceptual Identification of Tone Categories
  - Evaluation of Changes in Neural Tracking of Mandarin Lexical Tone F0 Patterns
  - Classification of FFRs to Mandarin Lexical Tones
  - Behavioral Statistical Analyses
  - EEG Statistical Analyses
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes one table and can be found with this article online at https://doi.org/10.1016/j.cub.2018.03.026.

ACKNOWLEDGMENTS

This work was supported by the National Institute On Deafness and Other Communication Disorders of the National Institutes of Health under award numbers R01DC015504 and R01DC013315 (B.C.). Earlier stages of this project were presented as podium presentations at the 2016 and 2017 mid-winter meetings of the Association for Research in Otolaryngology, where we received helpful feedback from peers. The authors would like to thank Jessica Roeder for the development of the dual-task and Erika Skoe for providing the MATLAB codes to create the autocorrelograms and implement the F0 tracking analysis. We also thank the members of the SoundBrain Laboratory for assistance with participant recruitment, data collection, and data preprocessing. Finally, the authors would like to thank three anonymous reviewers for their helpful comments and suggestions.

AUTHOR CONTRIBUTIONS

Conceptualization, B.C. and R.R.; Methodology, R.R., Z.X., and B.C.; Formal Analysis, R.R., Z.X., and F.L.; Investigation, R.R. and Z.X.; Resources, B.C.; Data Curation, R.R., Z.X., and F.L.; Writing – Original Draft, R.R. and B.C.; Writing – Review & Editing, R.R., Z.X., F.L., and B.C.; Visualization, R.R., Z.X., and F.L.; Project Administration, R.R. and B.C.; Funding Acquisition, B.C.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: September 23, 2017
Revised: January 17, 2018
Accepted: March 14, 2018
Published: April 19, 2018

REFERENCES

1. Zhang, Y., Kuhl, P.K., Imada, T., Iverson, P., Pruitt, J., Stevens, E.B., Kawakatsu, M., Tohkura, Y., and Nemoto, I. (2009). Neural signatures of phonetic learning in adulthood: a magnetoencephalography study. Neuroimage *46*, 226–240.
2. Wong, P.C., and Perrachione, T.K. (2007). Learning pitch patterns in lexical identification by native English-speaking adults. Appl. Psycholinguist. *28*, 565–585.
3. Werker, J.F., and Hensch, T.K. (2015). Critical periods in speech perception: new directions. Annu. Rev. Psychol. *66*, 173–196.
4. Kuhl, P.K., Tsao, F.-M., and Liu, H.-M. (2003). Foreign-language experience in infancy: effects of short-term exposure and social interaction on phonetic learning. Proc. Natl. Acad. Sci. USA *100*, 9096–9101.
5. Song, J.H., Skoe, E., Wong, P.C., and Kraus, N. (2008). Plasticity in the adult human auditory brainstem following short-term linguistic training. J. Cogn. Neurosci. *20*, 1892–1902.
6. Gold, J., Bennett, P.J., and Sekuler, A.B. (1999). Signal but not noise changes with perceptual learning. Nature *402*, 176–178.
7. Jurjut, O., Georgieva, P., Busse, L., and Katzner, S. (2017). Learning enhances sensory processing in mouse V1 before improving behavior. J. Neurosci. *37*, 6460–6474.
8. Ahissar, M., Nahum, M., Nelken, I., and Hochstein, S. (2009). Reverse hierarchies and sensory learning. Philos. Trans. R. Soc. Lond. B Biol. Sci. *364*, 285–299.
9. Hochstein, S., and Ahissar, M. (2002). View from the top: hierarchies and reverse hierarchies in the visual system. Neuron *36*, 791–804.
10. Pisoni, D.B., and Tash, J. (1974). Reaction times to comparisons within and across phonetic categories. Percept. Psychophys. *15*, 285–290.
11. Hallé, P.A., Chang, Y.-C., and Best, C.T. (2004). Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. J. Phonetics *32*, 395–421.
12. Bidelman, G.M., and Lee, C.-C. (2015). Effects of language experience and stimulus context on the neural organization and categorical perception of speech. Neuroimage *120*, 191–200.
13. Yi, H.-G., Maddox, W.T., Mumford, J.A., and Chandrasekaran, B. (2016). The Role of Corticostriatal Systems in Speech Category Learning. Cereb. Cortex *26*, 1409–1420.
14. Fahle, M. (2005). Perceptual learning: specificity versus generalization. Curr. Opin. Neurobiol. *15*, 154–160.
15. Skoe, E., and Kraus, N. (2010). Auditory brain stem response to complex sounds: a tutorial. Ear Hear. *31*, 302–324.
16. Coffey, E.B., Herholz, S.C., Chepesiuk, A.M., Baillet, S., and Zatorre, R.J. (2016). Cortical contributions to the auditory frequency-following response revealed by MEG. Nat. Commun. *7*, 11070.
17. Kraus, N., and White-Schwoch, T. (2015). Unraveling the Biology of Auditory Learning: A Cognitive-Sensorimotor-Reward Framework. Trends Cogn. Sci. *19*, 642–654.
18. Hélie, S., Waldschmidt, J.G., and Ashby, F.G. (2010). Automaticity in rule-based and information-integration categorization. Atten. Percept. Psychophys. *72*, 1013–1031.
19. Helie, S., Roeder, J.L., and Ashby, F.G. (2010). Evidence for cortical automaticity in rule-based categorization. J. Neurosci. *30*, 14225–14234.
20. Ashby, F.G., Ennis, J.M., and Spiering, B.J. (2007). A neurobiological theory of automaticity in perceptual categorization. Psychol. Rev. *114*, 632–656.
21. Carcagno, S., and Plack, C.J. (2011). Subcortical plasticity following perceptual learning in a pitch discrimination task. J. Assoc. Res. Otolaryngol. *12*, 89–100.
22. Wong, P.C., Skoe, E., Russo, N.M., Dees, T., and Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. Nat. Neurosci. *10*, 420–422.
23. Tierney, A.T., Krizman, J., and Kraus, N. (2015). Music training alters the course of adolescent auditory development. Proc. Natl. Acad. Sci. USA *112*, 10062–10067.

24. Xie, Z., Reetzke, R., and Chandrasekaran, B. (2017). Stability and plasticity in neural encoding of linguistically relevant pitch patterns. J. Neurophysiol. *117*, 1407–1422.

25. Watanabe, T., Náñez, J.E., Sr., Koyama, S., Mukai, I., Liederman, J., and Sasaki, Y. (2002). Greater plasticity in lower-level than higher-level visual motion processing in a passive perceptual learning task. Nat. Neurosci. *5*, 1003–1009.

26. Shamma, S. (2008). On the emergence and awareness of auditory objects. PLoS Biol. *6*, e155.

27. Ahissar, M., and Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. Trends Cogn. Sci. *8*, 457–464.

28. Fritz, J.B., David, S., and Shamma, S. (2013). Attention and dynamic, task-related receptive field plasticity in adult auditory cortex. In Neural correlates of auditory cognition (Springer), pp. 251–291.

29. Fritz, J.B., David, S.V., Radtke-Schuller, S., Yin, P., and Shamma, S.A. (2010). Adaptive, behaviorally gated, persistent encoding of task-relevant auditory information in ferret frontal cortex. Nat. Neurosci. *13*, 1011–1019.

30. Atiani, S., David, S.V., Elgueda, D., Locastro, M., Radtke-Schuller, S., Shamma, S.A., and Fritz, J.B. (2014). Emergent selectivity for task-relevant stimuli in higher-order auditory cortex. Neuron *82*, 486–499.

31. David, S.V., Fritz, J.B., and Shamma, S.A. (2012). Task reward structure shapes rapid receptive field plasticity in auditory cortex. Proc. Natl. Acad. Sci. USA *109*, 2144–2149.

32. Polley, D.B., Steinberg, E.E., and Merzenich, M.M. (2006). Perceptual learning directs auditory cortical map reorganization through top-down influences. J. Neurosci. *26*, 4970–4982.

33. Caras, M.L., and Sanes, D.H. (2017). Top-down modulation of sensory cortex gates perceptual learning. PNAS *114*, 9972–9977.

34. Goldstone, R.L. (1998). Perceptual learning. Annu. Rev. Psychol. *49*, 585–612.

35. Skoe, E., Chandrasekaran, B., Spitzer, E.R., Wong, P.C., and Kraus, N. (2014). Human brainstem plasticity: the interaction of stimulus probability and auditory learning. Neurobiol. Learn. Mem. *109*, 82–93.

36. Chandrasekaran, B., Kraus, N., and Wong, P.C. (2012). Human inferior colliculus activity relates to individual differences in spoken language learning. J. Neurophysiol. *107*, 1325–1336.

37. Lively, S.E., Logan, J.S., and Pisoni, D.B. (1993). Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. J. Acoust. Soc. Am. *94*, 1242–1255.

38. Bradlow, A.R., Akahane-Yamada, R., Pisoni, D.B., and Tohkura, Y. (1999). Training Japanese listeners to identify English /r/ and /l/: long-term retention of learning in perception and production. Percept. Psychophys. *61*, 977–985.

39. Roelfsema, P.R., van Ooyen, A., and Watanabe, T. (2010). Perceptual learning rules based on reinforcers and attention. Trends Cogn. Sci. *14*, 64–71.

40. Fritz, J., Shamma, S., Elhilali, M., and Klein, D. (2003). Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. Nat. Neurosci. *6*, 1216–1223.

41. Weinberger, N.M., Javid, R., and Lepan, B. (1993). Long-term retention of learning-induced receptive-field plasticity in the auditory cortex. Proc. Natl. Acad. Sci. USA *90*, 2394–2398.

42. Slee, S.J., and David, S.V. (2015). Rapid task-related plasticity of spectrotemporal receptive fields in the auditory midbrain. J. Neurosci. *35*, 13090–13102.

43. Weinberger, N.M. (2004). Specific long-term memory traces in primary auditory cortex. Nat. Rev. Neurosci. *5*, 279–290.

44. Bao, S., Chan, V.T., and Merzenich, M.M. (2001). Cortical remodelling induced by activity of ventral tegmental dopamine neurons. Nature *412*, 79–83.

45. Galván, V.V., and Weinberger, N.M. (2002). Long-term consolidation and retention of learning-induced tuning plasticity in the auditory cortex of the guinea pig. Neurobiol. Learn. Mem. *77*, 78–108.

46. Reed, A., Riley, J., Carraway, R., Carrasco, A., Perez, C., Jakkamsetti, V., and Kilgard, M.P. (2011). Cortical map plasticity improves learning but is not necessary for improved performance. Neuron *70*, 121–131.

47. Molina-Luna, K., Hertler, B., Buitrago, M.M., and Luft, A.R. (2008). Motor learning transiently changes cortical somatotopy. Neuroimage *40*, 1748–1754.

48. Wenger, E., Brozzoli, C., Lindenberger, U., and Lövdén, M. (2017). Expansion and Renormalization of Human Brain Structure During Skill Acquisition. Trends Cogn. Sci. *21*, 930–939.

49. Fu, M., and Zuo, Y. (2011). Experience-dependent structural plasticity in the cortex. Trends Neurosci. *34*, 177–187.

50. Holtmaat, A., and Svoboda, K. (2009). Experience-dependent structural synaptic plasticity in the mammalian brain. Nat. Rev. Neurosci. *10*, 647–658.

51. Suga, N. (2012). Tuning shifts of the auditory system by corticocortical and corticofugal projections and conditioning. Neurosci. Biobehav. Rev. *36*, 969–988.

52. Suga, N., Xiao, Z., Ma, X., and Ji, W. (2002). Plasticity and corticofugal modulation for hearing in adult animals. Neuron *36*, 9–18.

53. Krishnan, A., Gandour, J.T., and Bidelman, G.M. (2010). Brainstem pitch representation in native speakers of Mandarin is less susceptible to degradation of stimulus temporal regularity. Brain Res. *1313*, 124–133.

54. Llanos, F., Xie, Z., and Chandrasekaran, B. (2017). Hidden Markov modeling of frequency-following responses to Mandarin lexical tones. J. Neurosci. Methods *291*, 101–112.

55. Schneider, W., Eschman, A., and Zuccolotto, A. (2002). E-Prime: User's guide (Psychology Software Incorporated).

56. Elzhov, T.V., Mullen, K.M., and Bolker, B. (2009). minpack. lm: R Interface to the Levenberg-Marquardt Nonlinear Least-Squares Algorithm Found in MINPACK. R package version, 1.1-1.

57. Bates, D., Maechler, M., Bolker, B., and Walker, S. (2014). Package lme4: Linear mixed-effects models using Eigen and S4. R package version *67*.

58. Kuznetsova, A., Brockhoff, P.B., and Christensen, R.H.B. (2015). Package 'lmerTest'. R package version *2*.

59. Bidelman, G.M., Gandour, J.T., and Krishnan, A. (2011). Cross-domain effects of music and language experience on the representation of pitch in the human auditory brainstem. J. Cogn. Neurosci. *23*, 425–434.

60. Schön, D., Magne, C., and Besson, M. (2004). The music of speech: music training facilitates pitch processing in both music and language. Psychophysiology *41*, 341–349.

61. Chandrasekaran, B., Yi, H.-G., and Maddox, W.T. (2014). Dual-learning systems during speech category learning. Psychon. Bull. Rev. *21*, 488–495.

62. Shiffrin, R.M., and Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. Psychol. Rev. *84*, 127–190.

63. Xu, Y., Gandour, J.T., and Francis, A.L. (2006). Effects of language experience and stimulus complexity on the categorical perception of pitch direction. J. Acoust. Soc. Am. *120*, 1063–1074.

64. Skoe, E., Krizman, J., Spitzer, E., and Kraus, N. (2013). The auditory brainstem is a barometer of rapid auditory learning. Neuroscience *243*, 104–114.

65. Russo, N., Nicol, T., Musacchia, G., and Kraus, N. (2004). Brainstem responses to speech syllables. Clin. Neurophysiol. *115*, 2021–2030.

66. Chandrasekaran, B., and Kraus, N. (2010). The scalp-recorded brainstem response to speech: neural origins and plasticity. Psychophysiology *47*, 236–246.

67. Skoe, E., Krizman, J., Anderson, S., and Kraus, N. (2015). Stability and plasticity of auditory brainstem function across the lifespan. Cereb. Cortex *25*, 1415–1426.

68. Bidelman, G.M., Pousson, M., Dugas, C., and Fehrenbach, A. (2017). Test–retest reliability of dual-recorded brainstem versus cortical auditory-evoked potentials to speech. J Am Acad Audiol. *29*, 164–174.

69. Krishnan, A., Xu, Y., Gandour, J., and Cariani, P. (2005). Encoding of pitch in the human brainstem is sensitive to language experience. Brain Res. Cogn. Brain Res. *25*, 161–168.

70. Krishnan, A., Gandour, J.T., and Bidelman, G.M. (2012). Experience-dependent plasticity in pitch encoding: from brainstem to auditory cortex. Neuroreport *23*, 498–502.

71. Krishnan, A., Xu, Y., Gandour, J.T., and Cariani, P.A. (2004). Human frequency-following response: representation of pitch contours in Chinese tones. Hear. Res. *189*, 1–12.

72. Huang, W.-T., Liu, C., Dong, Q., and Nan, Y. (2015). Categorical perception of lexical tones in mandarin-speaking congenital amusics. Front. Psychol. *6*, 829.

73. Marquardt, D.W. (1963). An algorithm for least-squares estimation of nonlinear parameters. J. Soc. Ind. Appl. Math. *11*, 431–441.

74. Boersma, P. (1993). Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. In Proceedings of the Institute of Phonetic Sciences, *vol. 17* (University of Amsterdam), pp. 97–110.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited Data		
Behavioral data and stimuli	This paper; Mendeley Data	https://doi.org/10.17632/j2z2km4p9y.2
EEG data and stimuli	This paper; Mendeley Data	https://doi.org/10.17632/cvjjs4vdww.2
Software and Algorithms		
MATLAB	The MathWorks, Natick, MA	r2016a; https://www.mathworks.com
BrainVision Analyzer	Brain Products, Gilching, Germany	2.0; http://www.brainproducts.com
R, open source programming language	The R Core Team	3.4.1; https://www.r-project.org/
MATLAB toolbox for HMM modeling of FFRs	[54]	https://doi.org/10.1016/j.jneumeth.2017.08.010
E-Prime software	[55]	2.0.10
minpack.lm: R Interface to the Levenberg-Marquardt Nonlinear Least-Squares Algorithm	[56]	1.2-1
Lme4: Linear Mixed-Effects Models	[57]	1.1-13
lmerTest: Tests in Linear Mixed-Effects Models	[58]	2.0-33
Other		
BrainVision actiCHamp system	Brain Products, Gilching, Germany	EEG System; http://www.brainproducts.com

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources should be directed to and will be fulfilled by the Lead Contact, Bharath Chandrasekaran (bchandra@utexas.edu).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Twenty-two native English speaking adults (12 female; mean age = 20.9 year, SD = 4.1 year) were recruited to participate in the training study. Two participants did not complete all learning stages. One participant chose not to finish the training before criterion was met. The other participant did not demonstrate significant learning in the first 12 days of participation in the experiment, and therefore never progressed beyond the *novice* stage of learning. This participant was provided a choice to terminate or continue training and chose the former. Fifteen native Mandarin Chinese speaking adults (10 female, mean age = 24.0 year, SD = 2.7 year) were recruited to participate in a single day of behavioral testing to establish learning criterion for the English learners. We collected frequency-following responses (FFRs) to the four Mandarin tones (described below) and perceptual identification data from 13 of these participants. The FFR data from the native Chinese participants reported in this article was presented in a recent methods-related publication as part of the development of a novel machine learning algorithm to study the FFR [54]. Native English speaking participants reported that they were monolingual and had no previous exposure to or experience with a tonal language. All participants reported no current or previous history of a neurodevelopmental disorder or hearing deficit and no use of neuropsychiatric medication. Participants had normal hearing defined as air conduction thresholds < 20 dB HL at octave frequencies from 250 to 8,000 Hz measured by an Interacoustics Equinox 2.0 PC-Based Audiometer. Previous evidence has shown that music training influences the neural representation of non-native linguistic pitch patterns [22, 59, 60]. Therefore, participants with significant musical experience were excluded from participation in this study (all participants had < 6 years of continuous music training and were not currently practicing). We did not include a group of English participants that passively listened to Mandarin tones because previous evidence shows that extensive passive exposure to thousands of trials of Mandarin lexical tones over multiple days of recording does not modify participant neural tracking of Mandarin lexical tone F0 contours [24]. The Institutional Review Board at the University of Texas at Austin approved all materials and procedures, and all procedures were carried out following the approved guidelines. Written consent was obtained from all participants before participation in the study. All participants were recruited from the University of Texas at Austin community and received $15/hr monetary compensation for their participation.

METHOD DETAILS

Sound-to-Category Training Task

The sound-to-category training paradigm included high-talker variability stimuli and reinforcement via instructional feedback (Figure 1A), two training components found to direct participant attention to category-relevant acoustic cues [37–39], facilitating learning and long-lasting retention. The stimuli used in the speech training paradigm consisted of the four Mandarin lexical tones, which differ in their fundamental frequency (F0) contour: tone 1 (high-level), tone 2 (low-rising), tone 3 (low-dipping), and tone 4 (high-falling). Each tone was produced by two native Mandarin Chinese speakers (1 female), originally from Beijing, in the context of five syllables (*bu*, *di*, *lu*, *ma*, and *mi*). The 40 stimuli were normalized for root-mean-square (RMS) amplitude at 70 dB sound pressure level (SPL) and 440 ms duration. These stimuli were identical to the stimuli we have used in previous experiments [13, 24, 61].

Participants were instructed to categorize each stimulus into 1 of 4 categories by pressing the number keys (1, 2, 3 or 4) on a keyboard, corresponding to tone 1, tone 2, tone 3, and tone 4, respectively. No other instructions were provided. Each trial began with a fixation cross in the center of the screen for 750 ms. The stimulus was presented binaurally through Sennheiser HD 280 Pro circumaural headphones. After responding, participants were given feedback that was displayed for 1,000 ms. The response-to-feedback interval was fixed at 500 ms. The content of feedback was dependent on the accuracy of the response (''RIGHT'' versus ''WRONG''). Participants had unlimited time to respond, and the task moved on to the next trial once the participant input a response. Stimulus presentation, feedback, participant response, and reaction time (RT) measurement were controlled and acquired using MATLAB (The MathWorks, Natick, MA).

Dual-Task: Speech Categorization + Numerical Stroop

Participants simultaneously categorized Mandarin lexical tone stimuli while performing a numerical Stroop Task to test the emergence of 'automaticity' in performance [18, 19, 62]. Participants completed the dual-task during each learning phase (novice, experienced, over-trained) and at retention. The speech categorization stimuli in the dual-task were the same as those used in the sound-to-category training task. Participants were instructed that their goal was to remember which digit was larger in value, and which digit was larger in size. Each trial (80 trials in total) began with a fixation cross that appeared in the center of the screen for 750 ms. Two different digits (ranging from 2 to 8) were then randomly presented on the left and right sides of the screen at each trial as the tone stimulus was presented binaurally through Sennheiser HD 280 Pro circumaural headphones. One of the digits was displayed in a larger font relative to the other digit. At the end of each trial, participants would make a speech categorization response, as they did in the sound-to-category training and speech category generalization tasks. Before speech categorization, participants were cued either by the word ''Size'' or the word ''Value.'' If the cue was ''Size,'' the participant needed to indicate whether the digit of the larger size was on the right or the left of the fixation cross. If the cue was ''Value'' the participant needed to indicate whether the digit of the larger value was on the right or left of the fixation cross. Participants were instructed to focus on the new task and to perform the speech categorization task with the attentional resources they had left. Participants had unlimited time to respond and self-initiated to the next trial. Stimulus presentation, participant response, and RT measurement were controlled and acquired using MATLAB (The MathWorks, Natick, MA).

Generalization Task

The generalization task utilized an untrained set of 40 stimuli that were not used in the sound-to-category training task. The stimuli were derived from two untrained native Mandarin Chinese speakers (1 female), originally from Beijing, who produced the four Mandarin tones in citation form in the context of the same five syllables used in the training task. The 40 stimuli were normalized for RMS amplitude at 70 dB SPL and 440 ms duration. The trial procedure for the generalization task was consistent with the sound-to-category training task, with the exception that participants did not receive feedback after categorization responses. Participants had unlimited time to respond and self-initiated to the next trial. Stimulus presentation, participant response, and RT measurement were controlled and acquired using MATLAB (The MathWorks, Natick, MA).

Perceptual Identification Task

While tonal language speakers perceive discrete tone categories across a tone continuum, non-tonal language speakers do not [10–12, 63]. Compared to non-tonal language speakers, tonal language speakers demonstrate a steeper identification labeling curve (perceptual slope); and greater peak in identification reaction time (peak RT) at the category boundary, where between-category distinctions become ambiguous [10–12, 63]. To assess the training-induced changes in categorical perception of Mandarin lexical tone categories, participants completed a perceptual identification task during each learning phase (novice, experienced, over-trained) and at retention.

Stimuli consisted of a tone continuum identical to the continuum used in [63], where tone tokens were created to differ minimally acoustically but could be perceived categorically by native tonal language speakers [63]. The continuum was constructed by generating seven 300 ms tokens ranging in equal steps from the level to rising Mandarin lexical tone F0 contours. The F0 contours of the tone continuum were modeled by seven linear functions (see [63]). The resulting stimuli all had the same offset frequency (130.00 Hz), and therefore only differed in onset frequency [Step 1: 130.00 Hz; Step 2: 125.15 Hz; Step 3: 120.38 Hz; Step 4: 115.70 Hz; Step 5: 111.08 Hz; Step 6: 106.55 Hz; Step 7: 102.08 Hz]. The F0 contours of the stimuli are shown in Figure 1C.

Each stimulus was presented binaurally to participants via insert earphones (ER-3; Etymotic Research, Elk Grove Village, IL). Participants were instructed to press '1' if they heard a "level" pitch, or '2' if they heard a "rising" pitch. No feedback was provided to the participants. Each of the seven stimuli was randomly presented to each participant 20 times for a total of 140 trials. Participants had unlimited time to respond. After the participant made a response, the task moved on to the next trial following a 1000 ms delay.

Electrophysiology

Participants completed an electrophysiology session at each learning phase (novice, experienced, over-trained) and during retention. We recorded the frequency-following response (FFR), which is a sound-evoked response that mirrors the acoustic properties of the incoming acoustic signal with remarkable fidelity [5, 15, 64–67]. The FFR is considered an integrated response resulting from an interplay of early auditory subcortical and cortical systems [16, 17], shows high test-retest stability [15, 24, 67, 68], and robustly reflects long-term experience-dependent plasticity in native speakers of Mandarin [59, 69–71], as well as training-induced plasticity [5, 21, 22, 24, 64].

Stimuli

The stimuli consisted of four 250 ms synthetic Mandarin lexical tones minimally distinguished by their F0 contour (tone 1, tone 2, tone 3, and tone 4). The synthesis was derived from natural male production data. These stimuli were not used in the sound-to-category training, dual-task, and speech category generalization task. The four tones were superimposed over the same syllable /yi/, and only differed in their F0 contour: yi^1 high-level [tone 1], with F0 equal to 129 Hz; yi^2 low-rising [tone 2], with F0 rising from 109 to 133 Hz; yi^3 low-dipping [tone 3], with F0 onset falling from 103 to 89 Hz and F0 offset rising from 89 to 111 Hz; and yi^4 high-falling [tone 4], with falling F0 from 140 to 92 Hz. All tones were normalized to the same RMS amplitude at 72 dB SPL and duration at 250 ms.

Acquisition and Preprocessing

At each learning phase, participants sat in an acoustically attenuated booth and watched a muted movie or television show of their choice with subtitles. Electrophysiological responses to the Mandarin tone stimuli were collected using Ag-AgCl scalp electrodes, with the active electrode placed at the central zero (Cz) point, the reference at the right mastoid, and the ground at the left mastoid. Contact impedance was $< 5\ k\Omega$ for all electrodes for all recording sessions, and responses were recorded at a sampling rate of 25 kHz using Brain Vision PyCorder 1.0.7 (Brain Products, Gilching, Germany). Alternating polarities of the stimuli were binaurally presented via insert earphones (ER-3; Etymotic Research, Elk Grove Village, IL), with an inter-stimulus interval jittered between 122 to 148 ms. Consistent with previous studies, participants were instructed to ignore the sounds, focus on the selected movie or television show, and refrain from extraneous movement. The four Mandarin lexical tones were presented in separate blocks, and the order of blocks was counterbalanced across participants. Stimulus presentation was controlled by E-Prime 2.0.10 software [55].

The electrophysiological data were preprocessed with BrainVision Analyzer 2.0 (Brain Products, Gilching, Germany). Responses were offline bandpass filtered from 80 to 1,000 Hz (12 dB/octave, zero phase-shift). The bandpass filter approximately reflects the lower and upper limits of phase-locking along the auditory pathway that contributes to the FFR (auditory cortex, midbrain). Responses were then segmented into epochs of 310 ms (40 ms before stimulus onset and 20 ms after stimulus offset), and baseline corrected to the mean voltage of the noise floor (−40 to 0 ms). Epochs in which the amplitude exceeded ± 35 μV were considered artifacts and rejected. At each stage of learning, 1,000 artifact-free FFR trials (500 for each polarity) were obtained for each Mandarin lexical tone from all participants.

QUANTIFICATION AND STATISTICAL ANALYSIS

Evaluation of Changes in Perceptual Identification of Tone Categories

We evaluated the extent to which perceptual identification of tone categories changed as a function of training, by obtaining three well-established measures from each subject at each learning phase: category boundary, the slope of the identification curve, and the peak of identification reaction time (RT). To calculate the category boundary and the slope of the identification curve, we fitted a logistic regression model on the tone identification function on an individual subject basis, consistent with prior work [63, 72], with the following formula:

$$y = 1/(1 + exp(-b * (x - c)))$$

Where y refers to the proportion of participant responses that indicate "rising" pitch (ranging from 0 to 100%), x refers to the onset frequency of the stimulus' F0 contour, c refers to the category boundary where the proportion to report the tone as a "rising" pitch was 50%, and b refers to slope of the fitted logistic function and indicates the sharpness of the categorical boundary. Note that the tone identifications for "level" and "rising" responses are symmetrical, and therefore, only the "rising" responses were used in the current analysis. The model estimation procedures were conducted in R via the nlsLM function [56] that implemented the Levenberg-Marquardt algorithm [73] to search for the optimal parameters (b and c) that provide the best fit between the logistic model and the actual value y. Identification reaction times (RTs), also referred to as behavioral speech labeling speeds, were calculated as listeners' mean response latency across trials at each learning phase. In line with [12], RTs outside of 250-3500 ms were considered outliers and excluded from further analysis. To calculate the peak of identification RT, we estimated the difference of between-category perceptual sensitivity and within-category perceptual sensitivity [63]. Where between-category perceptual sensitivity was measured as identification RT from the categorical boundary (tone token 4) of the identification function; and within-category sensitivity was taken as the average identification RT near the ends of the tone continuum (tone tokens 2 and 6).

Evaluation of Changes in Neural Tracking of Mandarin Lexical Tone F0 Patterns

We evaluated the extent to which the FFRs follow F0 changes in the Mandarin lexical tone stimuli by extracting the F0 contour from the 1000-trial averaged FFRs using a periodicity detection short-term autocorrelation algorithm [74]. This algorithm works by sliding a 40-ms window over the time course of the FFR (10 to 260 ms post-stimulus onset). The 40-ms sliding window was shifted in 10 ms steps, to produce a total of 22 overlapping bins. The maximum (peak) autocorrelation value (ranging from −1 to 1) was searched over a lag value of 4 to 14.3 ms at each bin, a range that encompasses the time-variant periods of the F0 contours for the Mandarin tone stimuli. The peak autocorrelation value, as well as the corresponding lag, were recorded for each bin. The reciprocal of this time lag (or pitch period) was calculated to estimate the F0 for each bin. The resulting frequency values were concatenated to form a 22-point running F0 contour. The short-term autocorrelation algorithm was applied to both the FFRs and the Mandarin tone stimuli. Pitch tracking accuracy metrics were then computed using the F0 contour extracted from the FFRs and the F0 contour extracted from the stimuli.

Classification of FFRs to Mandarin Lexical Tones

Mandarin lexical tone categories were decoded from individual FFRs using the hidden Markov model (HMM) classifier [54]. The classifier was trained with sets of 500 FFRs per tone category. The remaining FFRs (500 per tone category) were used for testing. Training and testing sets were smoothed with a moving average of 200 FFRs. This combination of training, testing and averaging sizes provides optimal decoding of tone categories (and robust cross-language differences) in sets of 1000 FFRs [54]. The performance of the classifier was K-fold cross-validated. HMM, accuracy for each tone was computed from the cross-validated confusion matrix as the number of true positives and negatives over the number of true and false positives and negatives.

Behavioral Statistical Analyses

Linear mixed-effects regression (LMER) analyses were implemented on all behavioral variables to examine effects of training on behavioral performance (accuracy and median reaction time) in English learners, relative to native Mandarin performance. The native Mandarin performance was considered training criterion for reaching an expert level of speech categorization performance. Analyses were carried out in R, an open source programming language for statistical computing (R Development Core Team, 2014). We used the *lme4* package [57] and computed p values using the Satterthwaite's approximation for denominator degrees of freedom with the lmerTest package [58]. For all LMER models, we included one fixed-effect factor: training level (native, novice, experienced, over-trained, retention), with the performance of native Mandarin participants as the reference level. We additionally conducted an LMER analysis to investigate the interaction between training level (native, novice, over-trained, experienced, and retention) and tone token (1-7) on mean perceptual identification reaction time. In this analysis, the reference levels were native Mandarin performance and tone token 4, since native Mandarin performance was considered training criterion for other metrics, and tone token 4 is the categorical boundary of the tone continuum. The results of this analysis are reported in Table S1. All LMER models included by-participant random intercepts to account for inter-subject variability. All behavioral results reported are from a single model that included all training levels.

EEG Statistical Analyses

A two-way repeated-measures ANOVA was conducted to examine the effect of sound-to-category training on the neural tracking of Mandarin lexical tone F0 contours. In this analysis, learning phase (novice, experienced, over-trained, and retention) and stimulus (Tone 1, Tone 2, Tone 3, Tone 4) were included as within-subject factors. We examined two neural tracking metrics that have consistently demonstrated language experience-dependent plasticity, as well as training-induced plasticity: peak autocorrelation and stimulus-to-response correlation (for further details of metrics see:[24]). We report Greenhouse-Geisser corrected results for all ANOVA analyses.

DATA AND SOFTWARE AVAILABILITY

Our behavioral and EEG data are available to download via Mendeley Data at https://doi.org/10.17632/j2z2km4p9y.2 and https://doi.org/10.17632/cvjjs4vdww.2, respectively.