Systematic changes in neural selectivity reflect the acquired salience of category-diagnostic dimensions

Sahil Luthra[1], Raha N. Razin[2],
Adam T. Tierney[3], Lori L. Holt[4] and Frederic Dick[2*]

[1] Carnegie Mellon University, [2] University College London,
[3] Birkbeck, University of London, [4]The University of Texas at Austin

ORCID
Sahil Luthra: 0000-0002-3517-2609
Adam T. Tierney: 0000-0002-7624-6918
Lori L. Holt: 0000-0002-8732-4977
Frederic Dick: 0000-0002-2933-3912

Abstract

Humans and other animals develop remarkable behavioral specializations for identifying, differentiating, and acting on classes of ecologically important signals. Ultimately, this expertise is flexible enough to support diverse perceptual judgments: a voice, for example, simultaneously conveys what a talker says as well as myriad cues about her identity and state. Mature perception across complex signals thus involves both discovering and learning regularities that best inform diverse perceptual judgments, and weighting this information flexibly as task demands change. Here, we test whether this flexibility may involve endogenous attentional gain to task-relevant dimensions. We use two prospective auditory category learning tasks to relate a complex, entirely novel soundscape to four classes of "alien identity" and two classes of "alien size." Identity, but not size, categorization requires discovery and learning of patterned acoustic input situated in one of two simultaneous, frequency-delimited bands. This allows us to capitalize on the coarsely segregated frequency-band-specific channels of auditory tonotopic maps using fMRI to ask whether category-relevant perceptual information is prioritized relative to simultaneous, uninformative information. Among participants expert at alien identity categorization, we observe prioritization of the diagnostic frequency band that persists even when the diagnostic information becomes irrelevant in the size categorization task. Tellingly, the neural selectivity evoked implicitly in categorization aligns closely with activation driven by explicit, sustained selective attention to other sounds presented in the same frequency band. Additionally, we observe fingerprints of individual differences in the learning trajectories taken to achieve expert-level categorization in patterns of neural activity associated with the diagnostic dimension. In all, this indicates that acquiring categories can drive the emergence of acquired attentional salience to dimensions of acoustic input.

Systematic changes in neural selectivity reflect the acquired salience of category-diagnostic dimensions

Learners of all ages discover structured perceptual input to achieve their goals, and even simple sights and sounds convey rich information that can flexibly support multiple behaviors. A single utterance simultaneously conveys a talker's age (Lavan, 2022) and socioeconomic class (Kraus et al. 2019), whether she is a stranger or a loved one (Holmes et al., 2018), and if she is requesting her *bill* or her *pill* (Murphy et al. 2023). A glimpse of her face communicates her trustworthiness (Todorov et al. 2009), emotion (Krumhuber et al. 2023), and identity (Young et al. 2020) among other things. Successful perception thus requires that learners discover and learn distinct input patterns that best inform specific judgments. These underlying patterns are carried across multiple perceptual dimensions and – especially in the case of sound – often require integration of information across time. Flexibility in weighting the most diagnostic patterns of perceptual cues or dimensions for the task at hand is crucial for effective facial and vocal identification, object recognition and, at least in humans, speech comprehension.

Some have suggested that learning may drive endogenous attention to be directed toward diagnostic perceptual dimensions in a task-dependent manner (e.g., Gao et al. 2024; Nosofsky, 1986). These learned attentional biases weight perceptual dimensions most relevant to oft-performed tasks or behaviors (van Gulick & Gauthier, 2014; De Baene et al. 2008; O'Bryan et al. 2024) and may become so automatic that they influence neural activity even across passive exposure (Ley et al. 2012; Yin et al. 2020) or across an orthogonal task (Folstein et al. 2013). Dovetailing with results from studies of perceptual or cognitive expertise, particularly diagnostic perceptual dimensions may drive the character and intensity of neural activity to be similar across different stimulus domains due to common computational demands of the task at hand (Chua et al. 2015; McKeeff et al. 2010; Leech et al. 2009). In this way, dimensions that have historically been diagnostic may acquire salience across tasks.

For complex skills like spoken language and object recognition, it can take years or even decades to learn which perceptual dimensions are informative for a given behavioral goal and to optimally weight or attend to them (e.g., Idemaru & Holt, 2013; McMurray et al. 2018; Yurovsky et al. 2013; Yurovsky & Frank, 2017). Learners also progress along different trajectories to reach equivalent performance outcomes (Roark et al. 2024; Reetzke et al. 2018). Indeed, distinct perceptual and neural strategies are reflected in developmental neuroimaging studies demonstrating that there can be very different weightings of neural activation across brain regions even among children or adults who perform similarly on a task. These differences potentially reflect an individual's position along a long, and distinct, learning trajectory across which diagnostic input dimensions are discovered and tuned through experience.

Here, we prospectively introduce and explicitly manipulate a novel perceptual domain to establish whether *acquired, endogenous* attention directed toward diagnostic perceptual dimensions drives distinct patterns of auditory cortical activity, and how these patterns may differ among learners who take different trajectories to achieve equivalent expertise. Specifically, we train human adults in rapidly and accurately classifying sounds across a complex, multidimensional perceptual space composed of over 36,000 sounds varying in complex acoustic dimensions that are not readily verbalizable (Obasih et al., 2023). Both the sound exemplars and the categories we have defined across them are entirely novel to listeners, as are the 'space alien identity' and 'space craft size' to which they must map them according to spectrotemporal and perceived loudness cues, respectively.

All sounds possess acoustic patterns evolving in time across two simultaneous, non-overlapping frequency bands. Patterns diagnostic of *alien identity* are conveyed by one of the two bands and successful learning demands discovery of the diagnostic frequency band and learning across acoustically variable patterns evolving within it in the context of simultaneous, unstructured information in the other band. This aligns task demands with acoustic frequency, the primary axis of auditory representation, and allows us to measure cortical activation across the coarsely segregated frequency-band-specific channels of auditory tonotopic maps using fMRI. All sound exemplars also vary continuously in perceived loudness, with lower amplitude sounds diagnostic of 'small aliens' and higher amplitude sounds diagnostic of 'big aliens.' One group of participants trained over five days to learn alien identity (spectrotemporal dimension) and trained, as well, to categorize the same sound inventory according to alien size (amplitude dimension). Another group trained only on alien size (amplitude) thereby gaining no expertise with alien identity.

This allows us to answer several central questions. Are category-relevant perceptual dimensions prioritized relative to simultaneous uninformative dimensions? If so, does this selectivity persist when the diagnostic dimension is irrelevant to the task at hand? And do patterns of prioritization that emerge align closely with neural activity associated with explicit, sustained attention to the dimension across unrelated sounds? Finally, how do these patterns of neural activity relate to the distinct trajectories that learners take in reaching expert-level categorization?

Results

Category expertise develops across training, with generalization to novel exemplars

A group of young adults (N=95, see Methods for demographics) trained for 5 days to develop expertise in rapidly and accurately categorizing complex, multidimensional sounds they had never previously encountered. In the main 'alien identity' training task, four sound categories were associated with the identity of four distinct 'space alien' images (**Fig 1b**). Each sound possessed acoustic energy in two non-overlapping frequency bands (~80-750 Hz and ~1000-9000 Hz). Each band was populated with three 400-ms 'chirps' varying in frequency contour (**Fig 1a**). In the category-diagnostic band, these acoustically variable chirps possessed an underlying regularity that defined alien identity category membership. Simultaneous and temporally aligned chirps in the other, non-diagnostic frequency band were acoustically variable and possessed no coherent regularity. Category learning thus depended on discovering the diagnostic frequency band and learning the acoustically variable patterns within it. Two categories carried diagnostic information in the high-frequency band; two carried information in the low-frequency band (**Fig 1b**). This novel soundscape allowed us to capitalize on auditory cortical frequency sensitivity to establish cortical regions potentially relevant to the spectrally delimited category-diagnostic information signaling category identity.

Alien identity categorization training involved a two-alternative forced-choice (2AFC) task, with trials blocked by categories with high- vs. low-frequency category-diagnostic information. Trial-by-trial feedback indicated the correct alien category (see **Methods**; **Fig 1c**). After each training block, a four-alternative forced-choice (4AFC) task with no feedback assessed generalization of category learning across novel exemplars not experienced in training (**Fig 1c**). A separate training task involved categorizing exemplars drawn from all four alien identity categories as "big" or "small" according to stimulus amplitude, without regard to the spectrotemporal dimensions associated with alien identity (**Methods**; **Fig 1c**). This permits us to test whether stable changes in cortical activation associated with the frequency band carrying identity-diagnostic information persist when task demands are redirected to amplitude.

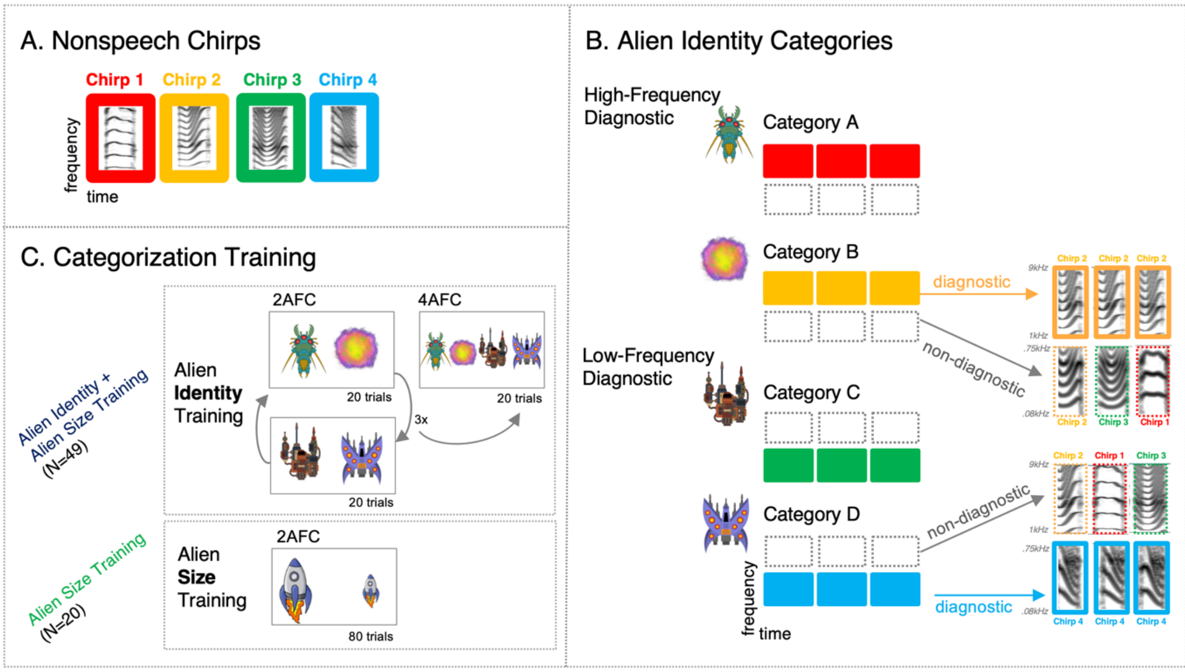


Figure 1. Stimuli. (A) Nonspeech Chirps. We extracted the fundamental frequency (F0) contour from natural speech recordings of single-syllable words varying in Mandarin lexical tone contour across four (2 female) native Mandarin speakers. This yielded 80 acoustically unique chirps from each of the four classes of Mandarin lexical tone contour (see Obasih et al. 2023). Spectrograms show one representative from each class (indicated by a different color). **(B) Alien Identity Categories.** We next shifted each nonspeech chirp into low- (~80-750 Hz) and high-frequency (~1000-9000 Hz) non-overlapping spectral bands. These spectrally shifted chirps served as building blocks for creating novel auditory category exemplars. Exemplars possessed simultaneous acoustic energy in each of the two spectral bands ((indicated by rectangles, color-coded by chirp class). For each category, one band conveyed category-diagnostic information (filled rectangles); the other, simultaneous, band (open dotted-line rectangles) did not. The category-diagnostic band was composed of a sequence of three unique chirps drawn from the same class of chirp such that chirps shared a common underlying structure (roughly: level, rising, dipping or falling; see **A**). For Categories A and B, the high-frequency band was diagnostic; for Categories C and D the low-frequency band was diagnostic. The non-diagnostic band was populated by three chirps drawn from different classes and thus conveyed no consistent structure. Chirps in the diagnostic and non-diagnostic bands were presented simultaneously. The insets on the right show a detailed example from Category B (high-frequency diagnostic, Chirp 2 structure) and Category D (low-frequency diagnostic, Chirp 4 structure). **(C) Categorization Training.** Across Days 1-5, one group of participants (N=49) learned to associate diverse exemplars drawn from the alien identity categories with one of four distinct "space aliens" via explicit feedback (see **B**, left). Two-alternative forced choice (2AFC) trials (blocked by high versus low diagnostic band, with feedback) were interleaved with four-alternative forced choice trials (no feedback) to assess generalization of learning. On Day 1 and Day 5, these participants learned to categorize exemplars drawn from all four alien identity categories as "big" or "small" aliens, according to exemplar amplitude (independent of alien identity). A separate group (N=20) trained only on this size judgment without training on alien identity.

A subset of expert participants (N=49) who achieved at least 75% 2AFC accuracy across both high- and low-frequency alien identity diagnostic bands by Day 5 returned an average of 9 days later for a single functional magnetic resonance imaging (fMRI) session. Overall, these experts exhibited above-chance 2AFC alien identity training performance on Day 1,

$t(48)=15.7$, $p = 7.35 \times 10^{-21}$, with improvements from Day 1 to Day 5, $t(48) = 7.82$, $p = 2.03 \times 10^{-10}$, to reach near-ceiling identity categorization (95.5%; SE=0.7%, **Fig 2a**). Among alien identity experts, generalization to novel exemplars exceeded chance on Day 1, $t(48)=5.64$, $p = 4.41 \times 10^{-7}$; M=68.5%, SE=3.3%), and improved significantly from Day 1 to Day 5 (M=87.7%, SE=1.9%, $t(48) = 6.56$, $p = 1.73 \times 10^{-8}$; **Fig 2b**). The trajectories across which learners reached criterion expertise was heterogeneous, variability that we used to subgroup learners for further analysis (**SI Fig S1**). This group of participants also trained on the alien size task on Days 1 and 5, with high performance (M=87.5%, SE=1.2%; **Fig 2c**).

In addition to the main group of participants expert in alien identity categorization, we also recruited, trained, and scanned a different set of listeners on *only* the alien size task (**Fig 1c**), with group size (N=20) matched to the highest-performing subgroup of the main identity-expert group (**SI Fig S1a**). These listeners achieved 87.1% (SE=1.7%) accuracy at the end of one session of alien size training (**Fig 2c**).

Both groups took part in a single fMRI session following training. In-scanner behavior echoed their patterns of training performance. Those who trained on both alien identity and alien size successfully categorized alien identity according to patterns evolving in the diagnostic frequency bands (2AFC task, M=95.8%, SE=0.7%; **Fig 2a**) and according to stimulus amplitude in the size task (M=86.2%, SE=1.1%; **Fig 2c**). Participants trained only on alien size performed equally well in training and during scanning (M = 84.2%, SE = 1.8%; **Fig 2c**).

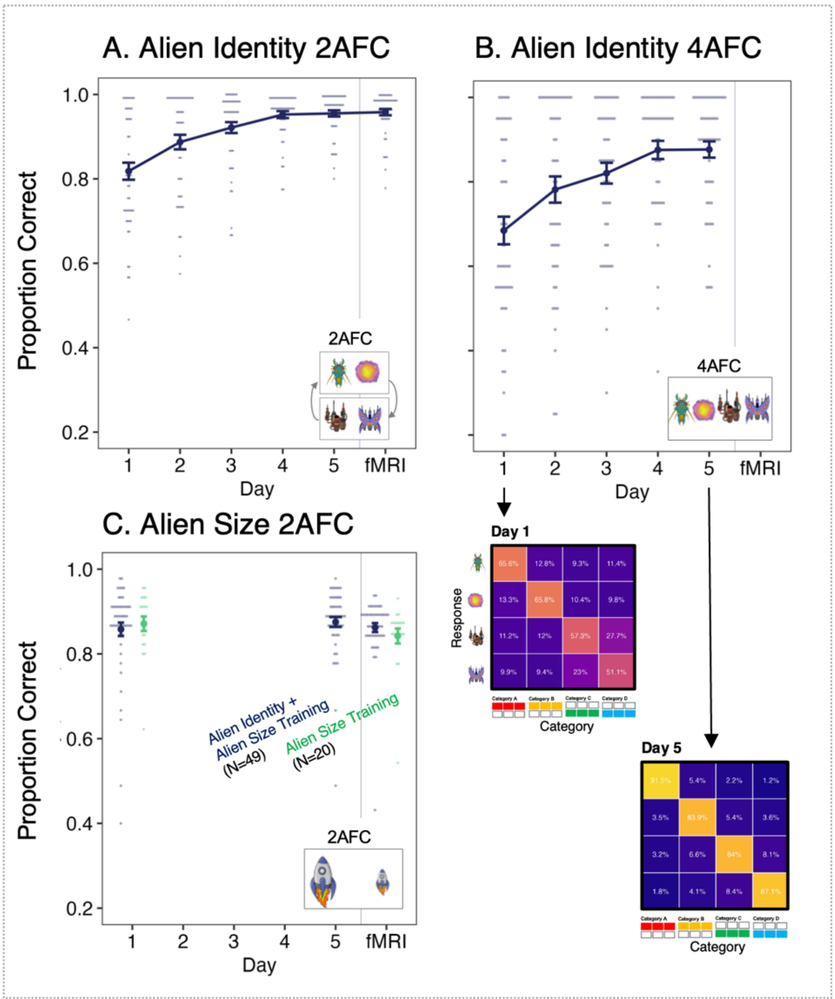


Figure 2. Behavioral Data. A. Alien Identity Training (2AFC with feedback). Proportion correct 2AFC categorization across Day 1-5 (and in the fMRI scanner, no feedback) among the 49 expert participants who achieved 75% accuracy across all alien identity categories by Day 5. **B. Alien Identity Generalization (4AFC no feedback).** The same 49 participants' 4AFC categorization of exemplars on which they were never trained to categorize according to alien identity improved across Days 1 to 5. Confusion matrices comparing category ground truth with response for Day 1 versus Day 5; Day 1 errors were predominantly within-frequency-band confusions. **C. Alien Size Training (2AFC with feedback).** A separate training task involved categorizing exemplars drawn from all four alien identity categories as "big" or "small" according to stimulus amplitude. The participants who trained on alien identity (N=49) trained on alien size on Day 1 and Day 5 of training, and in the scanner without feedback. An additional group of participants (N=20) trained only on alien size, never alien identity, for one day and performed the same task in the scanner with no feedback.

Cortical activation is greater for spectrally-selective category- regions, compared to non-diagnostic regions

We first present functional MRI data from the participant group trained on both the alien identity and size tasks. We characterize each participant's unique tonotopic organization on a voxel-wise basis across bilateral auditory-responsive cortex. This allows us to identify regions associated with category-diagnostic frequency bands, and to test differences in neural responses as a function of categorization task demands. Here, participants listened to concatenated series of 4-tone mini-sequences that stepped periodically over a 60-semitone range (175 to 5286 Hz, one run ascending and one descending, see Dick et al. 2017, and **Methods**). Participants performed a 1-back task on the mini-sequences, reporting infrequent repeats (in-scanner d': M=3.55, SE=0.08). From these data we compute voxel-wise activation to tone sequences grouped into the low and high frequency bands that conveyed alien identity information with the high-minus-low band difference in beta coefficients as a measure of spectral selectivity.

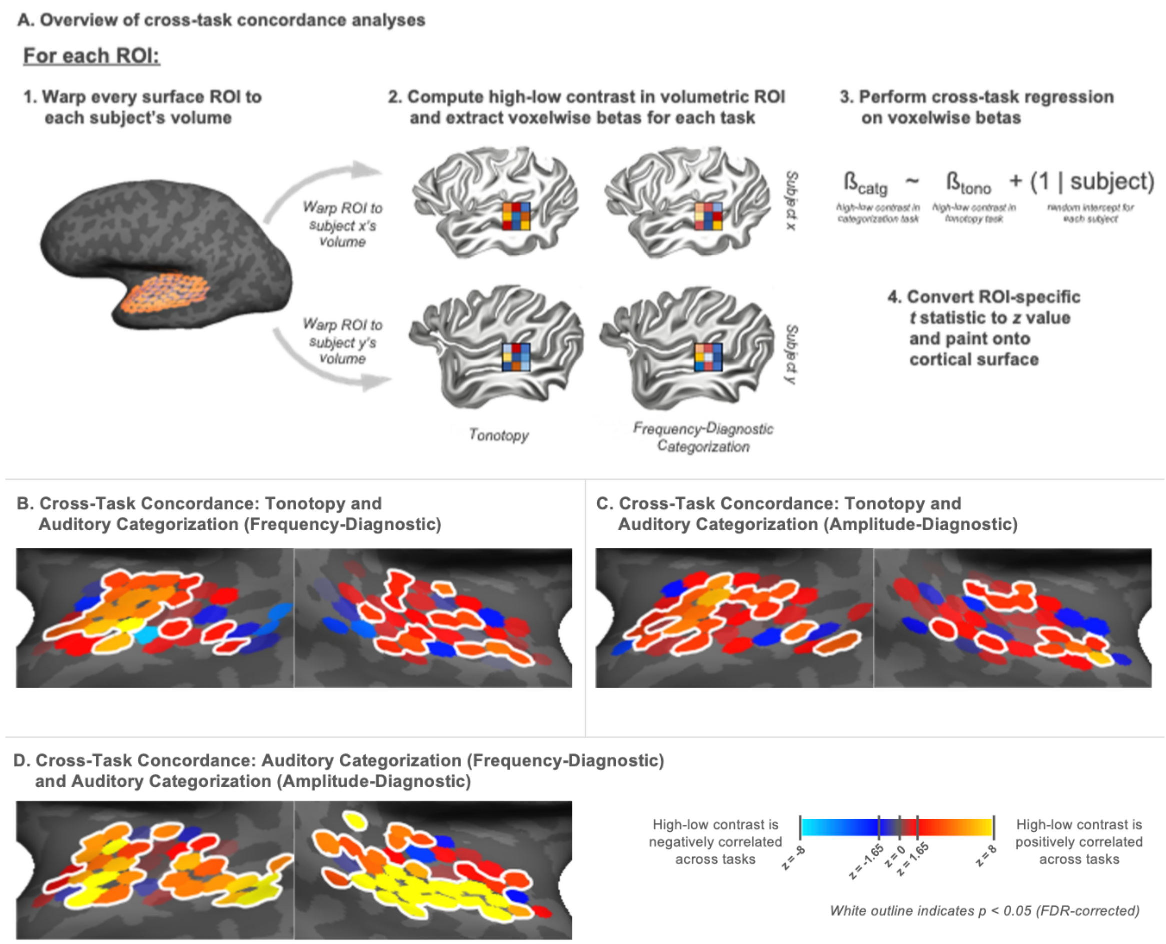


Figure 3. (A) To quantify the similarity in activation patterns across tasks, we defined a set of regions of interest (ROIs) and conducted cross-task concordance analyses within each ROI. First, we took a set of cortical surface ROIs that had been defined for a previous study (shown in orange; Dick et al., 2017); for the current work, we only considered ROIs that fell within tonotopically organized auditory cortex (purple outlines). Every ROI was then warped to each participant's native volumetric space. Within each volumetric ROI, we extracted voxel-by-voxel beta coefficients for the high-low contrast in each task. (For the two categorization tasks, we always examined the contrast between the high-band-diagnostic and the low-band-diagnostic categories; we used this contrast even when the task was to categorize stimuli based on overall amplitude, as we were always interested in testing for frequency-selective recruitment of auditory cortex. For the tonotopy task, we used the contrast between listening to high vs. low frequencies, and for the attention-o-tonotopy task, we used the contrast between attending to high vs. low frequencies.) To assess the similarity of activation across tasks, voxelwise betas were submitted to a regression analysis for each ROI separately; the goal in each regression analysis was to predict the beta coefficients from one task using the betas from another. Regressions also included by-subject random intercepts. The resultant ROI statistic, which indicates the strength and direction of the relationship between the two sets of beta coefficients, was painted on the cortical surface. **(B-D)** In this way, we assessed the concordance among the tonotopy, frequency-diagnostic categorization and amplitude-diagnostic categorization tasks. In general, there was a strong positive relationship among betas, indicated by warm-colored ROIs. ROIs outlined in white are statistically significant ($p < 0.05$ after applying a false discovery rate correction). Results indicate that during frequency-diagnostic categorization, there was greater cortical activation for category-diagnostic, compared to non-diagnostic, regions; strikingly, this pattern of activation persisted even when experts categorized stimuli along an orthogonal dimension (stimulus amplitude).

We observe coarse alignment of tonotopic organization across individuals (**SI Fig S2**), with cross-individual heterogeneity in the mosaic arrangement of fine-grained spectral selectivity, as expected from prior work (Moerel, de Martino, & Formisano, 2014). To capture consistent regional differences in responses across participants while retaining sensitivity at a voxel-wise level we adopt a region-of-interest (ROI) approach (see Dick et al. 2017). Each individual's auditory-responsive cortices are parcellated into a set of small ROIs defined over the cortical surface (**Fig 3**). Each ROI exhibiting spectral selectivity at the group level (as shown in the group-average tonotopic map) is warped and sampled into cortex to the individual's native volumetric EPI space. Voxel-by-voxel beta coefficients relating to activation during alien identity categorization (involving the high- and low-frequency category-diagnostic bands) are computed and a high-minus-low frequency-band difference in beta coefficients is calculated for each voxel, analogous to the high-minus-low spectral selectivity contrast from the tonotopy runs. These parallel difference scores permit us to fit a regression model across all participants within each ROI (with participant as a random factor) revealing the ROI-wise concordance between the voxelwise spectral selectivity evoked by tonotopy stimuli, and the hypothesized, category-specific frequency-band selectivity driven by the alien identity learning task (**Fig 3;** see Dick et al. 2017). In other words, this approach allows us to ask whether categorization decisions guided by chirp patterns embedded in high versus low frequency bands yield activation in corresponding spectrally selective cortical regions (**Fig 3**). Since each category exemplar possesses rich acoustic energy in each frequency band as well as amplitude variation (signaling size), an exaggerated neural response in the diagnostic versus non-diagnostic band is evidence of expertise-driven changes in cortical activation to the category exemplars.

We first established that these novel sounds evoked broad activation across much of auditory and auditory-related cortex bilaterally, as well as across a number of brain networks evident in the cortical-surface-based group average as well as for individual participants (**SI Fig S2**). We then turned to the ROI-based regression approach and found that, indeed, categorization decisions requiring high- versus low-frequency bands recruit regions that exhibit preferential responses to these frequencies. This pattern is especially strong in left anterior regions of auditory cortex, although there is also significant cross-task concordance in the right auditory cortex. The few ROIs with negative relationships across tasks tended to be situated in regions exhibiting relatively weak group-level tonotopic selectivity, suggesting that negative relationships may reflect cross-participant variability in tonotopic organization. Thus, we observed greater cortical activation within tonotopically mapped regions associated with the category-diagnostic frequency band, compared to the simultaneous non-diagnostic band, in the context of categorization.

Cortical activation reflects individuals' trajectories of alien identity category learning

The expert participants who completed the fMRI scan (N=49) had all achieved high levels of expertise in a novel domain. Their paths to expertise varied (**SI Fig S1**). To understand how these differences in learning trajectories might affect neural representations, we created subgroups of participants based on the speed and depth of learning over behavioral training. Early Experts (N=17) achieved at least 90% accuracy for both high- and low-band-diagnostic 2AFC training stimuli on Day 1 (**SI Fig S1c**). Late Experts (N=16) started at lower accuracy rates, but showed an accuracy increase of at least 20% over the course of 2AFC training (**SI Fig S1d**). We also estimated ultimate depth and generalizability of learning by separately creating groups based on generalization performance in the 4AFC task, with the Highest Achievers (N=16; **SI Fig S1a**) and Lowest Achievers (N=16; **SI Fig S1b**) defined across the top and bottom generalization terciles. These pairs of subgroups capture (partially overlapping) sets of participants with distinct learning trajectories.

We observe the fingerprints of these trajectories on subsequent cortical response to category exemplars. Compared to Late Experts, Early Experts exhibit stronger concordance of cortical activation within voxels showing a stimulus-evoked preference for the category-diagnostic frequency band, compared to the simultaneous non-diagnostic band, in the context of alien identity categorization (**Fig 4**). Similarly, the Highest Achievers exhibited greater concordance than the Lowest Achievers (**Fig 4**). These differences were especially pronounced in anterior left auditory cortex. This indicates that preferential activation for category-relevant frequency bands interacts with the degree of category expertise, and particularly with the rate of learning across training.

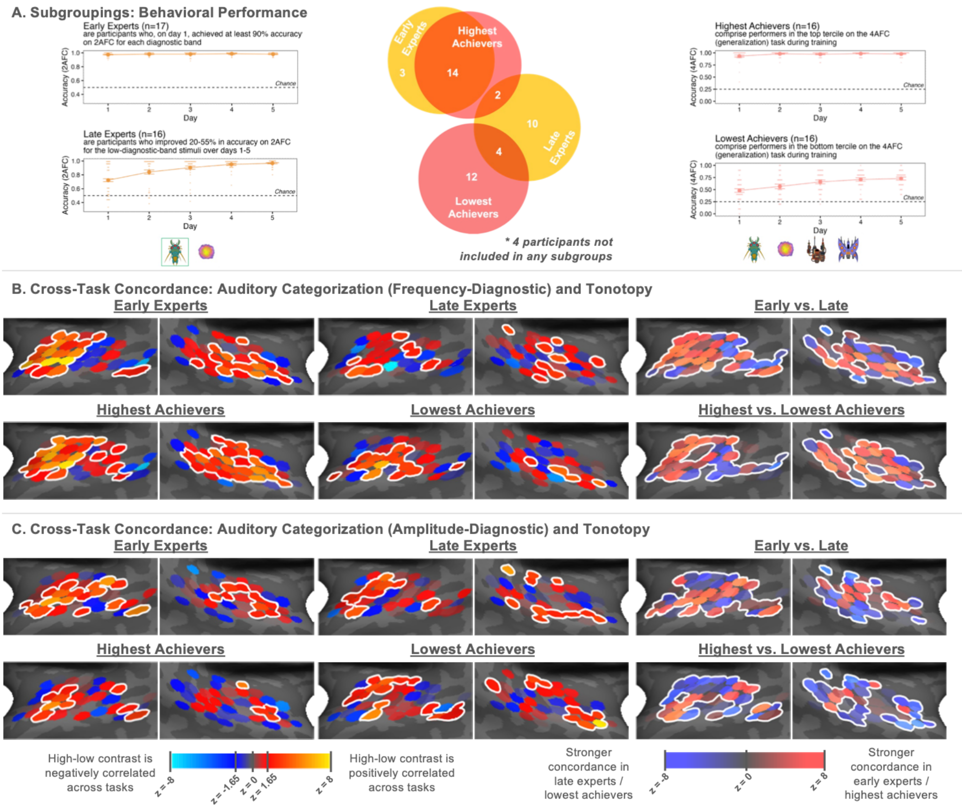


Figure 4. (A) Sub-group analyses tested for differences in cross-task concordance as a function of frequency-diagnostic categorization performance during the course of training. One set of analyses considered differences as a function of how quickly participants mastered the 2AFC categorization task: Early experts demonstrated mastery for both high-band-diagnostic and low-band-diagnostic categorization on day 1, whereas late experts demonstrated substantial improvement over the course of training. A second set of analyses considered differences as a function of generalization ability, separately analyzing data from the top tercile (highest achievers) and bottom tercile (lowest achievers) on the 4AFC generalization task. Behavioral performance and Venn diagrams for these sub-groups are shown. Light dots indicate individual participants, solid points indicate mean performance for each day, and error bars indicate standard error. **(B-C).** We assessed cross-task concordance between the tonotopy task and each of the two auditory categorization tasks. In each panel, the leftmost and center sets of images show concordance for each group separately. The rightmost set of images illustrates the difference in concordance between groups; warm colors indicate ROIs where stronger cross-task concordance was seen in early experts / highest achievers, and cool colors indicate ROIs where stronger cross-task concordance was seen in late experts / lowest achievers.

Greater cortical activation for category-diagnostic regions persists even in contexts that do not demand this information

We hypothesize that training caused identity-diagnostic dimensions to acquire salience, such that frequency-selective recruitment persists even when task demands shift to categorization across an orthogonal dimension. To test this, participants also categorized exemplars from all four alien categories in different blocks according to whether they were “big” or “small” aliens according to perceived loudness, not frequency. As noted above, in-scanner categorization was accurate (but well below ceiling performance), indicating that participants successfully relied upon the amplitude dimension for categorization, and also that the task was non-trivial. Although task demands directed participants' overt attention to an orthogonal acoustic dimension (amplitude), concordance between voxels' stimulus-evoked frequency preference and exemplars’ frequency-diagnostic band for the over-practiced and more challenging *identity* task persisted.

Moreover, there was strong concordance between the differences in activation evoked by high- and low-diagnostic band exemplars during *identity (*frequency-diagnostic) categorization task blocks and *size* (amplitude-diagnostic) categorization blocks (**Fig 3d**). The differential cortical activation across high- and low-frequency bands cannot easily be attributed to 'bottom-up' acoustic salience, as all stimuli possessed information in each band in addition to amplitude variation. Rather, this is consistent with an ‘acquired salience’ for category-diagnostic dimensions that reflect stable changes in cortical activation elicited by category exemplars persistent across changing task demands; prioritization of category-diagnostic dimensions in auditory areas is present even in contexts that do not demand reliance on them.

The degree of acquired salience of diagnostic dimension depends on learning trajectory and categorization expertise

As described above, expert-level categorization drove activation of auditory areas tuned to acoustic frequencies that were diagnostic of alien identity, even when experts were successfully categorizing stimuli based on stimulus amplitude. To make a critical test of the importance of learning to weight or attend to the identity-diagnostic spectrally spectral band in driving these results, we compared activation in our Early Expert group to a that of a group of participants (N=20) trained only on the alien size task (based on stimulus amplitude). These participants remained naïve to alien identity and *never* trained to learn categories based on frequency-band-delimited acoustic information.

Alien-size-trained participants exhibited a roughly equivalent degree of spectral selectivity and tonotopic organization in auditory cortex as the group that trained on both alien identity and size (**Fig 5**), with high behavioral accuracy on the tonotopic mapping task (in-scanner d': M=2.68, SE=0.13). This allowed us to test whether, like alien-identity-trained participants, alien-size-only-trained participants' auditory cortex would show differential activation for stimuli with identity-disambiguating information in high and low spectral bands when they were performing the amplitude-based alien size task. Though they demonstrated high behavioral accuracy on amplitude-based categorization (M: 84.2%, SE: 1.8%) they exhibited little sign of activation for the identity-diagnostic frequency band that matched underlying frequency preferences (**Fig 5**). By comparison, the Early Experts trained on both identity and size tasks showed strong concordance across ROIs between the difference in activation to high- versus low-diagnostic frequency band stimuli, and high- versus low-frequency tone mapping during alien-size categorization blocks. Most ROIs showed significantly greater spectrally-selective modulation among participants trained on both alien identity and size, compared to those trained on alien size only (**Fig 5**). This shows that learning categories dependent upon patterns of information evolving in frequency-selective bands drove the acquired salience of the perceptual dimension, such that experts in identity categorization (but not identity-naïve participants) exhibited frequency-specific recruitment of auditory cortex, even when categorizing stimuli along an orthogonal dimension (stimulus amplitude).

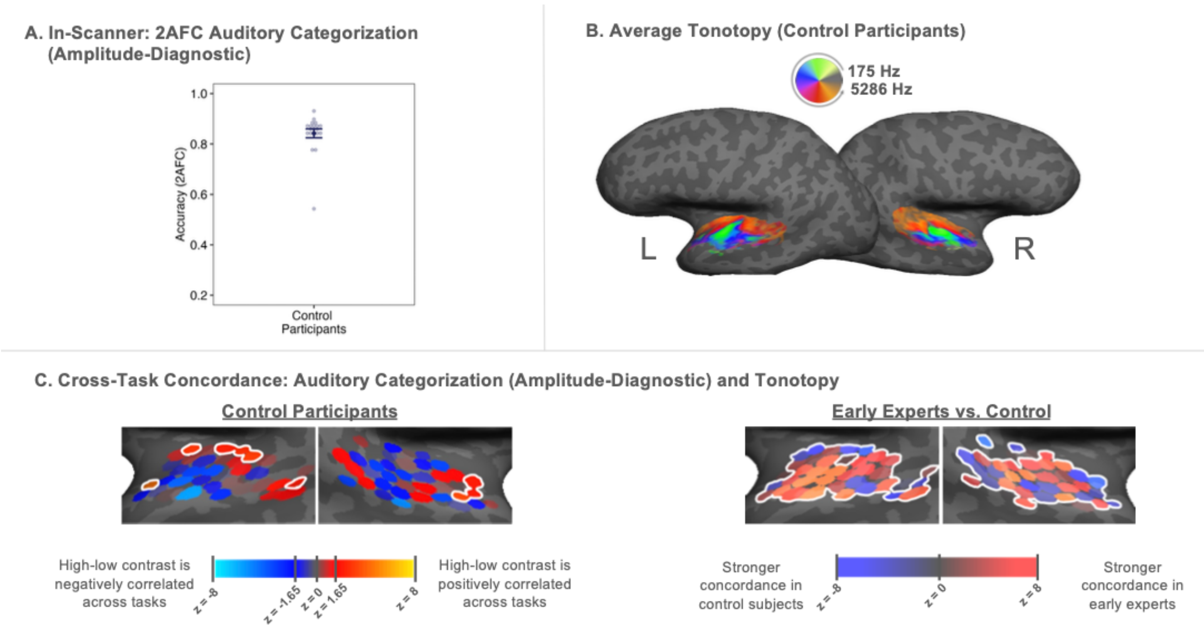


Figure 5. A group of Alien Size trained participants (n=20) were trained to categorize stimuli based on overall amplitude but critically never learned to use spectrally-delimited information for determining category identity. **(A)** These participants showed high in-scanner accuracy on amplitude-based categorization and (**B**) exhibited tonotopic selectivity in auditory cortex bilaterally. **(C)** Crucially, however, these participants did not recruit auditory cortex in a frequency-selective way when performing amplitude-based categorization, in contrast to our sample of experts. This suggests that the frequency-selective recruitment of auditory cortex that experts exhibited in the amplitude categorization task was not driven by properties of the acoustic signal *per se*. Rather, training on frequency-based auditory categorization drove the acquired salience of the frequency dimension, such that experts (but not control participants) exhibited frequency-specific recruitment of auditory cortex, even when categorizing along an orthogonal dimension (amplitude).

Spectrally-selective activation patterns driven by category learning and overt spectral attention align

We observe that when categorization implicitly hinges on information in a delimited frequency band, voxels more responsive to this frequency band show increased activation compared to when diagnostic information is in the less-preferred frequency band, despite simultaneous acoustic energy in each. We next asked whether the prioritization of category-diagnostic information is consistent with hypotheses that changes in perceptual weighting reflect attentional gain. Among the full group of participants trained on alien identity, we compared identity-categorization-associated activation differences with differences evoked by explicit sustained attention to one of two simultaneously presented streams of sinewave tones situated within the high- and low-frequency bands.

In addition to being scanned during tonotopy and alien categorization tasks, the identity-trained participants also completed a sustained auditory selective attention task based on the approach of Dick et al. (2017) who found that overt attention directed at sequences of tones positioned in specific frequency bands elicits an attention-driven map across auditory-responsive cortex. Participants heard two simultaneously presented sequences of sinewave tones composed of concatenated 4-tone mini-sequences that varied in amplitude; mini-sequences were played simultaneously in higher- and lower-frequency bands, corresponding to the diagnostic bands of the alien identity task. At the beginning of each task block, explicit instructions directed participants to attend to detecting mini-sequence repeats in the high-frequency band, the low-frequency-band, or repeats of the overall relative amplitude of the mini-sequence (1-back task d': M=2.28, SE=0.13).

As anticipated given previous studies of spectrally-selective attention (Da Costa et al., 2013; de Martino et al. 2015; Dick et al. 2017; Riecke et al. 2017), voxelwise differences in activation when attending to high- versus low-frequency bands were strongly associated with stimulus-evoked frequency preference (from tonotopy scans) across most auditory cortex ROIs. This was particularly true in lateral ROIs and in the right hemisphere.

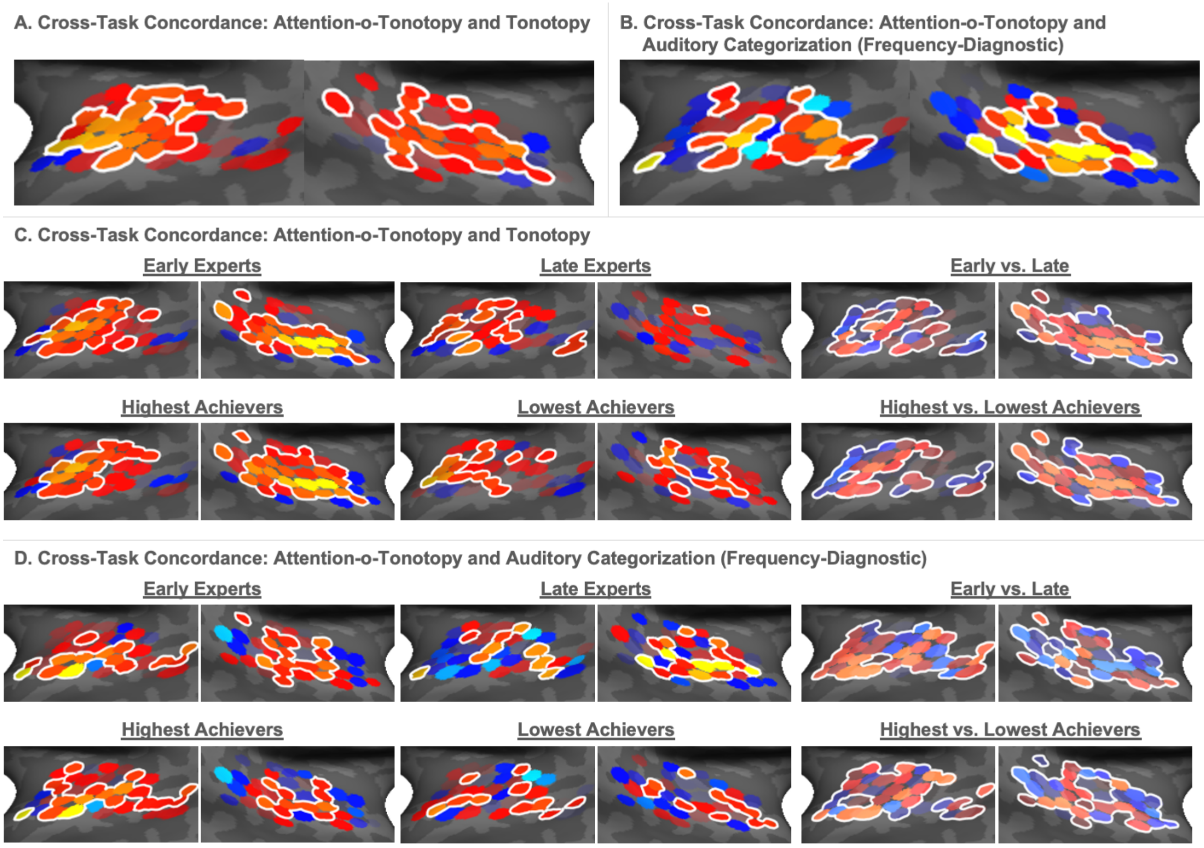


Figure 6. Theoretical accounts suggest that the neural selectivity that emerges with learning is driven by increased attention to diagnostic dimensions. To test this hypothesis, we assessed cross-task concordance with the attention-o-tonotopy task, where participants were directed to explicitly attend to particular frequency bands. Analyses considering all participants showed strong concordance between attention-driven activation and **(A)** the activation elicited by the tonotopy task (i.e., stimulus-driven activation), as well with **(B)** categorization-driven activation. **(C-D)** Sub-group analyses showed generally stronger concordance in early experts compared to late experts, as well as in the highest achievers compared to lowest achievers (the latter assessed with the 4AFC generalization task).

We then asked whether patterns of cortical activation elicited by explicit, directed attention to the frequency bands align with activation patterns that arise in category decisions implicitly reliant on information in these same frequency bands. Cross-task concordance maps showed generally strong activation alignment (**Fig 6**); the few ROIs that exhibited negative correlations tended to be those for which tonotopic selectivity was weaker or those along sharp transitions between high-frequency-preferring and low-frequency-preferring regions. Thus, expert categorization elicited activation in regions preferentially recruited when listeners explicitly attend to tones situated in the category-diagnostic region of the frequency dimension. Across multiple auditory cortical regions, activation patterns elicited by explicit, sustained attention aligns with the diagnostic-band prioritization that arose implicitly in categorization. In other words, regions recruited when listeners explicitly attend to a delimited region of the frequency spectrum are correspondingly more active when categorization *implicitly* hinges on information situated in these frequency bands.

Of additional note, Early Experts and High Achievers showed stronger concordance between attention-driven and stimulus-driven tonotopic maps (**Fig 6**) as well as stronger similarity in the activation profiles elicited by the overt attention and frequency-diagnostic auditory categorization tasks (**Fig 6**). Thus, even among proficient experts, greater expertise is associated with stronger alignment between stimulus-driven, attention-driven, and categorization-driven maps of frequency-selective neural activity in auditory cortex.

Discussion

How does perception come to weight the most diagnostic patterns of input for the task at hand? We used fMRI to probe the neural consequences of learning to categorize across a novel soundscape defined by multiple, complex acoustic dimensions evolving in time. Situating category-diagnostic patterns into spectrally delimited frequency bands allowed us to capitalize on coarsely segregated frequency-band-specific channels of auditory tonotopic maps to make targeted predictions of how learning-driven modulation of cortical activity would intersect with existing tonotopic functional regionalization, how different trajectories of learning to expert-level performance would be reflected in these representations, and whether prioritization of category-diagnostic information aligns with overt attention directed toward distinct stimuli in the same frequency bands.

We observe that category learning drives prioritization across a category-relevant dimension, as reflected in greater cortical activation within tonotopically mapped regions associated with the category-diagnostic frequency band, compared to the simultaneous non-diagnostic band. This prioritization persists even in contexts that do not require categorization according to the diagnostic dimension and is impacted by the trajectory of learning participants take to become expert and is absent among category-naïve listeners. Further, it aligns closely with the patterns of cortical activation that emerge when listeners explicitly direct sustained attention to unrelated sounds situated in the same frequency band, consistent with ‘attentional gain’ accounts of advantaged perceptual processing of category-diagnostic dimensions. In all, learning appears to drive the emergence of acquired attentional salience to category-relevant perceptual dimensions.

The question of whether, and if so how, sensory cortical representations are impacted by becoming expert in a categorization domain has not had a clear answer. Evidence of cortical modulation with category expertise comes predominantly from visual cortex under conditions of active categorization (Sigala and Logothetis, 2002; De Baene et al. 2008) but with mixed results. Some studies do not observe effects of categorization on visual cortex (Freedman et al. 2003; Jiang et al. 2007; Gillebert et al. 2008; van der Linden et al. 2010), potentially indicative of category representations arising in flexible, amodal cortical areas with little lasting influence on sensory cortex (Freedman et al. 2003; Serre et al. 2007; Roy et al. 2010) with sensory cortical effects arising from top-down modulation in the context of active categorization. In contrast, our results indicate stable, lasting learning-driven change in auditory cortex that persists even when attention is not directed toward the category-diagnostic dimension (see also Folstein et al. 2013 and Van Gulick & Gauthier, 2014 for evidence from visual categories).

Indeed, Yin et al. 2020 present evidence for both top-down modulation of auditory cortex by frontal cortex and also persistent, intrinsic auditory cortical patterns related to acquired categories evident even in passive listening. Though we limited our investigation specifically to tonotopically mappable temporal cortex, these auditory regions receive rich top-down feedback inputs and exhibit modulation by behavioral context, extending their response beyond basic representation of perceptual dimensions (e.g., Zhong et al. 2019). This invites the possibility of top-down categorization-driven modulation -- well-documented in nonhuman animal electrophysiology and, particularly in the visual system (e.g., Freedman et al. 2002; DeGutis and D'Esposito, 2007) – influences that interact with the persistent learning-driven changes to auditory cortical representation outside of active categorization that we observe.

Whether this modulation of neural activity is a consequence of learned 'attentional gain' to diagnostic dimensions also has surprisingly little direct support, despite the centrality of this proposition in theories of visual and speech categorization. It is clear that explicit attention to stimulus features modulates representations in both visual (Foster & Ling, 2022; Gundlach et al., 2023; Saenz et al. 2002; Serences & Boynton, 2007; Treue & Maunsell, 1999; Yoo et al., 2022) and auditory (DaCosta et al. 2013; Dick et al. 2017; Reicke et al. 2017) cortical response. Here, we asked whether these patterns of cortical gain are concordant with patterns that arise implicitly across category-diagnostic dimensions in categorization. Consistent our prior research (Dick et al., 2017), sustained auditory selective attention across sequences of tones evolving simultaneously across the same two frequency bands results in greater activation across the attended frequency band. Here, in the same listeners, these patterns aligned closely with the prioritization of frequency-selective cortical activity driven by categorization demands. Of note, this concordance is not driven by task demands; prioritization of the diagnostic frequency band persists even when task demands change and frequency-band is no longer task-relevant. This suggests that prioritization of diagnostic information has become a stable, lasting part of how category exemplars among experts that is absent in category-naïve listeners. This presents the first empirical evidence linking acquired salience to category-diagnostic dimensions of acoustic input and overt, directed selective attention.

Methods

Participants

Participants from the Pittsburgh, PA, USA and London, UK communities completed a five-day training protocol (N=95; 18-40 yrs). All were native-English speakers with normal hearing, had normal or corrected-to-normal vision and no history of neurological impairment or speech/language disorder, and reported no experience with a tonal language. Individuals who met criterion levels of expertise (75% across both diagnostic frequency bands on the 2AFC alien identity task on Day 5; see "Behavioral Training," below) were invited to participate in a subsequent MRI session, with data collected from N=54 experts. Five participants' data were excluded (motion, non-compliance with task) resulting a final sample of N=49 experts (18-37 yrs, M=23.9 yrs; 29 female, 18 male, 2 non-binary; 59% White; 32 Pittsburgh, 17 London). As described below (see "Alien Size Only group"), we also recruited a sample of non-experts who participated in an abbreviated online training and MRI session (N=20, 18-43 yrs, M=28.1 yrs; 12 female, 8 male; 55% White; 20 London).

Procedure

Online training across 5 daily sessions (60-90 min; see **Fig 1**) was implemented across Gorilla software (Anwyl-Irvine, Massonnié, Flitton, Kirkham, & Evershed, 2020), the Google Chrome browser, wired headphones (with compliance screening as in Milne et al., 2021), and a computer/laptop (not tablet/phone). Stimuli were created using custom code in Matlab version 2021b (The MathWorks), Praat (Boersma, 2001) and SoX version 14.4.2 (www.sourceforge.net).

Behavioral Training. Participants learned to associate 1400 ms novel sounds with one of four alien images via category-informative feedback identifying the correct alien after each response. The four auditory categories were sampled from a highly complex perceptual space (see Obasih et al., 2023). In brief, each stimulus was composed of patterned acoustic energy in two non-overlapping acoustic frequency bands (**Fig 1;** ~ 80-750 Hz and ~1000-9000 Hz). Each band was populated with three 400-ms tonal 'chirps' (100-ms ISI) varying in frequency contour (**Fig 1**). In the *alien-identity-diagnostic* frequency band (high-frequency for Categories A, B and low-frequency for Categories C, D), the 3 acoustically variable chirps possessed an underlying regularity that defined alien identity category membership (**Fig 1**). Temporally aligned chirps in the non-diagnostic frequency band (low for Categories A, B; high for Categories C, D) were acoustically variable and possessed no coherent regularity.

To create a large set of naturally varying category exemplars, we derived chirps from 4 native Mandarin Chinese talkers (2 female) uttering single-syllable words in each of Mandarin's four lexical tones, which convey variable, but structured, fundamental frequency (F0) contours. We extracted the F0 contour from each, and resynthesized the contour to create a non-speech chirp, time-normalizing each to be 400 ms (**Fig 1**). We shifted each chirp +33 semitones then applied a 1000 Hz high-pass filter to create a pool of chirps to populate a high-frequency band; equivalently we applied a -1 semitone plus a 500-Hz low-pass filter to create constituents for the low-frequency band. Three distinct chirps derived from the same talker and the same Mandarin lexical tone constituted the category-diagnostic band, conveying a regularity in frequency contour. Three different chirps were derived from three different Mandarin tones (from any of the four categories) to populate the non-diagnostic frequency band such that there was no regularity across chirps. In all, this created a stimulus pool of 36,000 exemplars (Obasih et al., 2023) from which we sampled 592 unique stimuli. (Full details in *SI Text*).

In addition to the *Alien-Identity-Category*-associated spectral information, we also imposed orthogonal amplitude variation (between 62-80 dB SPL) on copies of each exemplar. This had two goals: It increased variability and task difficulty, and also permitted creation of two *Alien-Size categories,* with "Big Alien" exemplars falling between 72-80 dB sound intensity (all reported dB referenced to $2*10^{-5}$ and measured in Praat), and "Small Alien" exemplars falling between 62-70 dB. Thus, each exemplar could be categorized along Identity (spectral) or Size (amplitude) dimensions.

For Alien Identity training, each of the 5 daily sessions involved four 120-trial 2AFC alien identity categorization training blocks where participants indicated identity with a keypress, and feedback (1500 ms) indicated the correct alien. Trials were blocked according to the high- or low-frequency category-diagnostic band (Categories A vs. B and Categories C vs. D) such that every 20 trials, the category pair, and thus diagnostic frequency band, alternated; the other two alien response options were blanked out as response alternatives. Training stimuli were randomly selected from a pool of 512 exemplars (128/category). Amplitude level was balanced across trials and roughly equated across categories on each day. Each training block was followed by a 20-trial 4AFC categorization task with all aliens as response options, and no feedback. These trials involved a separate pool of 20 exemplars/category that were never presented in training to assess generalization of learning (see **Fig 1**). Across both 2AFC and

4AFC tasks, category exemplars were mixed with a fixed level of scanner noise (57 dB) to to familiarize participants with categorization in an MRI environment. The visual position of aliens (and associated keypress) was fixed the first two days of training, then allowed to vary starting on Day 3. (We should note that Obasih et al. 2023 found that alien category learning speed and attainment was surprisingly unaffected by different training regimes, including manipulations of exemplar variability, blocking versus interleaving categories in training, and overt instructions to attend to the diagnostic band),

On Day 1 and Day 5, participants also trained on the 'Alien Size' category size task. Across the first eight trials, one exemplar from each of the four categories was presented at the highest amplitude (80 dB), then one exemplar from each alien category at the lowest amplitude (62 dB). The amplitude difference diminished each successive 4 trials for a total of 40 trials, with 71 dB defining the big vs. small alien category boundary. In a second 40-trial block, amplitude values were fully randomized with equiprobable big/small exemplars and an equal number of exemplars from each alien category. A fixed level of scanner noise (again 57 dB) was mixed with stimuli to mimic the MRI sound environment.

Tonotopy. To characterize voxel-wise frequency selectivity efficiently for each participant and create tonotopic maps, while being scanned, participants listened to sequences of 178-ms sinewave tones (0 ms ISI, 5 ms ramp at on/offset) organized in 4-tone mini-sequences, each separated by 356 ms. Participants reported infrequent mini-sequence repeats in a 1-back task. The frequency range of the four tones composing the mini-sequences increased or decreased in frequency incrementally in 10 logarithmically scaled steps to sweep across 175-5286 Hz (60 semitones). Across the 60 mini-sequences contained in each 64 sec sweep, there were 3 mini-sequence repeats that occurred quasi-randomly. The swept frequency permitted both fMRI analysis using phase-encoded mapping, where voxels responsive to a particular frequency should respond at a consistent phase delay (Dick et al., 2012; 2017; Sereno et al., 1995) as well as multiple regression analysis to identify voxels responsive to the range of the low-frequency category-diagnostic frequency band (< 500 Hz, lowest three steps) and the high-frequency category-diagnostic frequency band (≥ 1000 Hz, highest five steps).

For analysis of training and in-scanner behavior, the response window for each mini-sequence was defined from the onset of the final tone to the onset of the final tone in the subsequent sequence (1068-ms response window). For calculation of d-prime scores, extreme hit and false alarm rates of 0 or 1 were adjusted following the approach of Stanislaw and Todorov (1999). As an introduction to the task, participants familiarized with one up-sweep and one down-sweep on Day 4 of online training, with mini-sequence repeats indicated on-screen. As with the space alien task, during out-of-scanner training, scanner EPI noise (57 dB) was mixed in with tone stimuli (~80 dB).

Attention-o-tonotopy. To characterize voxelwise spectrally-selective and amplitude-selective overt attentional responses in individual participants, and following the approach of Dick et al. (2017), participants heard 4-tone mini-sequences evolving simultaneously in two frequency bands with instructions to “attend low” or “attend high” and report mini-sequence repeats from the attended frequency band. Here, an additional condition directed participants to “attend volume,” reporting 1-back matches on the attended amplitude dimension. Mini-sequences were composed of 180-ms sinewave tones drawn from a pool of four frequencies (1, 3, 5 and 7 semitones above a ‘base’ frequency). Tones were drawn with replacement, with the constraint that a sequence could never be a single tone played four times. Simultaneous tones from high- and low-frequency bands were temporally aligned. Mini-sequences were separated by 360 ms silence in short blocks that concluded with 2 sec of silence. Each block included two mini-sequence repeats in the high-frequency band, two repeats in low-frequency band, and two amplitude repeats. The 1080-ms response window for each mini-sequence was defined from the onset of the final tone to the onset of the final tone in the subsequent sequence.

Participants familiarized with the task on Day 3 of online training, and first experienced mini-sequences in single frequency bands of fixed amplitude. Next, amplitude variability was introduced, then finally, participants practiced with dual-frequency-band amplitude-variable stimuli. On-screen messages alerted participants to mini-sequence repeats in the attended band. There was no scanner noise mixed with stimuli during familiarization. Participants then completed 24 attention-o-tonotopy trials (3 blocks/trial, 11-17 mini-sequences/block), with on-screen feedback provided for the first half of the trials. For half of the trials, the dual-band stimuli had base frequencies of 280 and 1253 Hz, and for the other half, the base frequencies were 400 and 2123 Hz. “Attend high,” “attend low” and “attend volume” trials were equiprobable. The amplitude of the stimuli was set so that the two frequency bands were approximately equated on perceptual loudness (~73 dB average), and a fixed level of scanner noise (57 dB SPL) was presented throughout these trials.

MRI Procedure. Images were acquired in a single 90-min session with 3T Siemens Prisma scanners and 32-channel head coils (CMU-Pitt BRIDGE Center in Pittsburgh, RRID:SCR_023356, and at the Birkbeck-UCL Centre for Neuroimaging in London). Head movement among London participants was minimized with a head stabilization prototype (MR Minimal Motion System). Both sites delivered audio over Sensimetrics earbuds with pre-filtering to accommodate the earbuds’ response profile, presented tasks using PsychoPy (v. 2022.2.4), and recorded responses with a button box. Identical pulse sequences were used across sites, and very similar QA procedures were in place by the respective MRI center teams to monitor scanner performance.

Experts qualifying for the MRI session completed a 15-min online refresher one day prior to the scan. This involved two tonotopy trials (one up-sweep, one down-sweep), 80 2AFC categorization trials (48 alien identity, 32 alien size, drawn from the pool of 80 generalization exemplars with no feedback), and two short 9-block attention-o-tonotopy trials with parameters mirroring the MRI task.

At the scanning session, structural images were acquired using a T1-weighted magnetization-prepared rapid acquisition gradient echo (MPRAGE) sequence (TR = 2300 ms, TE = 2.98 ms, FOV = 256 mm, flip angle = 9°) with 1 mm sagittal slices. Functional echo planar images were acquired using a T2*-weighted sequence (TR = 1.0 s, TE = 30 ms, 44 slices, 2.0 mm thickness, in-plane resolution = 2 mm × 2 mm with 6/8 partial Fourier encoding, FOV = 212 mm, flip angle = 62°, multiband acceleration factor = 4); we included eight dummy volumes at the start of each scan to allow the scanner to reach B1 equilibrium. Functional images were acquired with anterior-to-posterior phase encoding; following each task, we collected two additional volumes with reversed phase encoding to correct for image distortion caused by B0 inhomogeneities.

The scanning session started with MPRAGE acquisition. Then, participants completed four runs of the tonotopy task (4 sweeps/run; 1st and 3rd up-sweep, 2nd and 4th down-sweep). Participants fixated on a central cross overlaid on a small landscape image to minimize sweep-related eye movements; images changed over the course each run (5 images/run; random duration to avoid aliasing with tonotopic sweep rate).

After tonotopy runs, there were six runs (10 blocks, 8 trials/block) of 2AFC categorization. 20 blocks involved categorization of alien identity across the high-frequency diagnostic band, 20 involved categorization of identity across the low-frequency band, and 20 involved size categorization across amplitude. A pseudorandomized block order common across participants assured that blocks reliant on the same diagnostic dimension did not occur consecutively. Overall block order in runs 1-3 repeated in runs 4-6, but with trial order shuffled, and with the amplitude set to a different level within the 'big' or 'small' amplitude range. In all, there were 320 alien identity trials (80/category, with equivalent exemplars from the two size categories) and 160 alien size trials (with equal exemplars from the four identity categories).

To facilitate comparisons between the alien identity and alien size blocks, 20 of the 40 blocks that appeared during the alien identity task also appeared in the alien size task, with the same fixed trial order within the block. Half of the alien size blocks appeared before their corresponding alien identity blocks, and half appeared after. Each block began with a visual display of the task (“Alien Identity” or “Alien Size,” 3 sec). Each trial was 3 seconds in length (1.5 seconds for audio playout, 1.5-second response window).

Finally, there were two runs of the attention-o-tonotopy task (one with base frequencies of 280 and 1253 Hz, one with base frequencies of 400 and 2123 Hz). Each run involved 15 blocks (22-34 sequences per block), with 12-15 sec of rest after every 3 blocks (except after the final block in a run). Auditory instructions at the beginning of each block indicated the locus of attention (“high,” “low,” “volume,” “rest”). As in the tonotopy task, participants fixated on a central cross overlaid on changing landscape images (5 images/run).

Alien Size Only Group. An additional participant group completed an abbreviated scanner session, which included structural imaging, four runs of the tonotopy task, and two runs of 2AFC categorization. Critically, the 2AFC categorization task only involved alien size trials, and the control group received the same alien size trials (in the same order) as the expert group. Prior to the scanner session, these participants completed one session of Alien Size training and the tonotopy training; the training procedure for these tasks was identical to the procedure for the expert participants.

MRI Analyses. Cortical surfaces were reconstructed for each participant from the T1-weighted MPRAGE using FreeSurfer (Dale & Sereno, 1993; Dale, Fischl, & Sereno, 1999; Fischl, 2012).

Functional images from the tonotopy task were minimally preprocessed using AFNI (Cox, 1996) to account for saturation by discarding the first 8 volumes, to unwarp images using phase-reversed images collected at the end of each task, and to align all volumes to a reference volume from the middle of the first run. Next, the reference volume was aligned to the cortical surface using boundary-based registration (Greve & Fischl, 2009) in FreeSurfer.

To establish the regions of auditory cortex showing spectral selectivity, we calculated tonotopic maps using Fourier-based analyses on the preprocessed functional data from the tonotopy task in csurf, following standard phase-encoded mapping approaches (Sereno et al., 1995; Dick et al. 2012, 2017). Relying on the fact that each frequency step occurs at a consistent time within a sweep, Fourier analysis allowed us to compute a set of F-statistics indicating how each voxel responds at the frequency of stimulus cycling (4 cycles per run, 4 64-sec sweeps per run) relative to other frequencies; the phase lag (i.e., the delay relative to the start of the cycle) of the maximal response can be used to determine the frequency preference of each voxel. We performed Fourier analysis of each run (up-sweeps time-reversed for phase averaging with down-sweeps).

Finally, we painted the phase of the signal (indicating frequency preference) as a color map and visualized the data on each participant’s cortical surface reconstruction. Projecting each subject's data onto an icosahedral surface and averaging across participants created a group map that was projected to a single participant's surface reconstruction for visualization. Subsequent region-of-interest (ROI) analyses, described below, were limited to regions in auditory cortex where tonotopic organization was observed at the group level in expert participants.

We next performed a regression-based analysis on the preprocessed functional data from each of the three tasks using 3dREMLfit in AFNI (Chen et al., 2012). For each analysis, regressors were constructed by convolving a vector of the onset times for each frequency step with a square wave of the appropriate duration. (Rigid-body motion parameters (movement in the x-, y- and z-axis directions, as well as pitch, roll, and yaw) were included as regressors of no interest). The resultant beta estimates were used to compute, for each task, the difference in activation between the high-frequency condition and low-frequency condition. In the categorization tasks, this corresponds to the difference between performing categorization on high-frequency-diagnostic stimuli vs. low-frequency-diagnostic stimuli. In the attention-o-tonotopy task, this corresponds to the difference between explicitly attending to high or low frequency mini-sequences. Results were mapped to the cortical surface using the mri_vol2surf Freesurfer command. To generate group-level contrast maps, subject-level maps were projected onto an icosahedral mesh and averaged, the result of which was displayed on a single subject's surface.

Although there is a macroscopic cross-participant similarity in large-scale tonotopic gradients, there is considerable cross-participant variability in frequency preference at a finer grain, particularly when voxels are aligned by cortical folding patterns (Besle et al., 2018; Moerel et al., 2014). Thus, to measure regional encoding and cross-task correspondence in spectral attentional weighting while taking this variability into account, we calculated the degree of correspondence in cross-task voxel-wise spectral selectivity in a set of small cortical surface ROIs defined in a previous study (Dick et al., 2017). Here, we limit our analysis to those ROIs where tonotopic selectivity was observed at the group level in our expert participants. For each participant, the ROIs that tessellate tonotopically-organized auditory cortex (37 per hemisphere) were projected from a single participant to the unit icosahedron (sphere_reg), then projected again to each curvature-aligned participant's surface, and finally to the bbregister-aligned native space 2 x 2 x 2 mm EPI space using Freesurfer's mri_label2vol.

For each ROI, cross-task concordance was evaluated via regression analysis, where all subject-wise beta coefficients in one task (e.g., $b_{\text{tonotopy(high-low)}}$) was used to predict the high-low betas in another task (e.g., $b_{\text{attention-o-tonotopy(high-low)}}$); each regression model included random intercepts for each participant. The resultant *t* values were transformed to *z* statistics, and these *z* values were painted onto the cortical surface. Significance (alpha < 0.05) was assessed using one-tailed tests, as we had *a priori* hypothesized a positive relationship between betas across tasks. To control for multiple comparisons, we applied a false discovery rate (FDR) correction on our *p* values using the 3dFDR command in AFNI. Note that because 3dFDR performs a two-tailed adjustment, *p* values associated with negative *z* statistics were inverted (i.e., 1-p) both prior to and following FDR correction, yielding FDR-corrected values that reflect one-tailed tests. Similar analyses were performed for different subgroups of participants defined based on behavioral performance during training.

We also conducted a series of analyses to compare cortical responses across groups (e.g., in participants who acquired category expertise early in training vs. those who acquired expertise over the course of training). We tested whether the resultant cross-task concordance statistics differed across groups by computing a $z_{\text{difference}}$ statistic (Hays, 1994), defined as:

$$z_{difference} = \frac{z_{group1} - z_{group2}}{\sqrt{\frac{1}{n_{group1} - 3} + \frac{1}{n_{group2} - 3}}}$$

These values were projected onto a single subject's cortical surface, providing an index of which group demonstrated stronger cross-task concordance. For these analyses, significance was assessed using two-tailed tests and FDR corrections were applied to control for multiple comparisons.

Data Availability

The datasets and code generated and analyzed in the current study are available at https://osf.io/arjcf/.

References

Besle, J., Mougin, O., Sánchez-Panchuelo, R.-M., Lanting, C., Gowland, P., Bowtell, R., … Krumbholz, K. (2018). Is Human Auditory Cortex Organization Compatible With the Monkey Model? Contrary Evidence From Ultra-High-Field Functional and Structural MRI. *Cerebral Cortex*, 28(Suppl 1), 287–19. https://doi.org/10.1093/cercor/bhy267

Chua, K. W., Richler, J. J., & Gauthier, I. (2015). Holistic processing from learned attention to parts. *Journal of Experimental Psychology: General*, *144*(4), 723–729. https://doi.org/10.1037/xge0000063

Costa, S. D., Zwaag, W. V. D., Miller, L. M., Clarke, S., & Saenz, M. (2013). Tuning in to sound: frequency-selective attentional filter in human primary auditory cortex. *Journal of Neuroscience*, 33(5), 1858–1863. Retrieved from https://doi.org/10.1523/jneurosci.4405-12.2013

Da Costa S, van der Zwaag W, Marques JP, Frackowiak RS, Clarke S, Saenz M (2011) Human primary auditory cortex follows the shape of Heschl's gyrus. J Neurosci 31:14067–14075. 10.1523/JNEUROSCI.2000-11.2011

De Baene, W., Ons, B., Wagemans, J., & Vogels, R. (2008). Effects of category learning on the stimulus selectivity of macaque inferior temporal neurons. *Learning & Memory*, *15*(9), 717–727. https://doi.org/10.1101/lm.1040508

De Baene, W., Ons, B., Wagemans, J., & Vogels, R. (2008). Effects of category learning on the stimulus selectivity of macaque inferior temporal neurons. *Learning & Memory*, *15*(9), 717–727. https://doi.org/10.1101/lm.1040508

DeGutis, J., & D'Esposito, M. (2007). Distinct mechanisms in visual category learning. *Cognitive, affective & behavioral neuroscience*, *7*(3), 251–259. https://doi.org/10.3758/cabn.7.3.251

Dick, F. K., Lehet, M. I., Callaghan, M. F., Keller, T. A., Sereno, M. I., & Holt, L. L. (2017). Extensive Tonotopic Mapping across Auditory Cortex Is Recapitulated by Spectrally Directed Attention and Systematically Related to Cortical Myeloarchitecture. The Journal of neuroscience : the official journal of the Society for Neuroscience, 37(50), 12187–12201. https://doi.org/10.1523/JNEUROSCI.1436-17.2017

Folstein, J. R., Palmeri, T. J., & Gauthier, I. (2013). Category learning increases discriminability of relevant object dimensions in visual cortex. *Cerebral Cortex (New York, N.Y.: 1991)*, *23*(4), 814–823. https://doi.org/10.1093/cercor/bhs067

Folstein, J. R., Palmeri, T. J., & Gauthier, I. (2013). Category learning increases discriminability of relevant object dimensions in visual cortex. *Cerebral Cortex*, *23*(4), 814–823. https://doi.org/10.1093/cercor/bhs067

Foster, J. J., & Ling, S. (2022). Feature-Based Attention Multiplicatively Scales the fMRI-BOLD Contrast-Response Function. *The Journal of Neuroscience*, *42*(36), 6894–6906. https://doi.org/10.1523/JNEUROSCI.0513-22.2022

Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2003). A comparison of primate prefrontal and inferior temporal cortices during visual categorization. Journal of Neuroscience, 23, 5235–5246.

Gao, M., Turner, B. M., & Sloutsky, V. M. (2024). The Role of Attention in Category Representation. *Cognitive Science*, *48*(4), e13438. https://doi.org/10.1111/cogs.13438

Gillebert, C. R., Op de Beeck, H. P., Panis, S., & Wagemans, J. (2008). Subordinate categorization enhances the neural selectivity in human object-selective cortex for fine shape differences. Journal of Cognitive Neuroscience, 21, 1054–1064.

Gundlach, C., Wehle, S., & Müller, M. M. (2023). Early sensory gain control is dominated by obligatory and global feature-based attention in top-down shifts of combined spatial and feature-based attention. *Cerebral cortex (New York, N.Y. : 1991)*, *33*(19), 10286–10302. https://doi.org/10.1093/cercor/bhad282

Hays, W. L. (William L. (1994). *Statistics / William L. Hays* (5th ed). Harcourt Brace College Publishers.

Holmes, E., Domingo, Y., & Johnsrude, I. S. (2018). Familiar Voices Are More Intelligible, Even if They Are Not Recognized as Familiar. *Psychological Science*, *29*(10), 1575–1583. https://doi.org/10.1177/0956797618779083

Idemaru, K., & Holt, L. L. (2013). The developmental trajectory of children's perception and production of English /r/-/l/. *The Journal of the Acoustical Society of America*, *133*(6), 4232–4246. https://doi.org/10.1121/1.4802905

Jiang, X., Bradley, E., Rini, R. A., Zeffiro, T., Vanmeter, J., & Riesenhuber, M. (2007). Categorization training results in shape- and category-selective human neural plasticity. Neuron, 53, 891–903.

Kraus, M. W., Torrez, B., Park, J. W., & Ghayebi, F. (2019). Evidence for the reproduction of social class in brief speech. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(46), 22998–23003. https://doi.org/10.1073/pnas.1900500116

Krumhuber, E.G., Skora, L.I., Hill, H.C.H. *et al.* The role of facial movements in emotion recognition. *Nature Reviews Psycholology,* **2**, 283–296 (2023). https://doi.org/10.1038/s44159-023-00172-1

Lavan N. (2022). The effect of familiarity on within-person age judgements from voices. *British Journal of Psychology, 113*(1), 287–299. https://doi.org/10.1111/bjop.12526

Leech, R., Holt, L. L., Devlin, J. T., & Dick, F. (2009). Expertise with artificial nonspeech sounds recruits speech-sensitive cortical regions. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *29*(16), 5234–5239. https://doi.org/10.1523/JNEUROSCI.5758-08.2009

Ley, A., Vroomen, J., Hausfeld, L., Valente, G., De Weerd, P., & Formisano, E. (2012). Learning of new sound categories shapes neural response patterns in human auditory cortex. *The Journal of Neuroscience*, 32(38), 13273–13280. https://doi.org/10.1523/JNEUROSCI.0584-12.2012

Liu, S. T., Montes-Lourido, P., Wang, X., & Sadagopan, S. (2019). Optimal features for auditory categorization. *Nature Communications*, *10*(1), 1302. https://doi.org/10.1038/s41467-019-09115-y

Martino, F. D., Moerel, M., Ugurbil, K., Goebel, R., Yacoub, E., & Formisano, E. (2015). Frequency preference and attention effects across cortical depths in the human primary auditory cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 112(52), 16036–16041. Retrieved from https://doi.org/10.1073/pnas.1507552112

McKeeff, T. J., McGugin, R. W., Tong, F., & Gauthier, I. (2010). Expertise increases the functional overlap between face and object perception. *Cognition*, *117*(3), 355–360. https://doi.org/10.1016/j.cognition.2010.09.002

McMurray, B., Danelz, A., Rigler, H., & Seedorff, M. (2018). Speech categorization develops slowly through adolescence. *Developmental Psychology*, *54*(8), 1472–1491. https://doi.org/10.1037/dev0000542

Moerel, M., Martino, F. D., & Formisano, E. (2014). An anatomical and functional topography of human auditory cortical areas. *Frontiers in Neuroscience*, 8. Retrieved from https://doi.org/10.3389/fnins.2014.00225

Murphy, T. K., Nozari, N., & Holt, L. L. (2024). Transfer of statistical learning from passive speech perception to speech production. *Psychonomic Bulletin & Review*, *31*(3), 1193–1205. https://doi.org/10.3758/s13423-023-02399-8

Nosofsky R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, *115*(1), 39–61. https://doi.org/10.1037//0096-3445.115.1.39

O'Bryan, S. R., Jung, S., Mohan, A. J., & Scolari, M. (2024). Category Learning Selectively Enhances Representations of Boundary-Adjacent Exemplars in Early Visual Cortex. *The Journal of Neuroscience*, *44*(3), e1039232023. https://doi.org/10.1523/JNEUROSCI.1039-23.2023

Obasih, C. O., Luthra, S., Dick, F., & Holt, L. L. (2023). Auditory category learning is robust across training regimes. *Cognition*, *237*, 105467. https://doi.org/10.1016/j.cognition.2023.105467

Reetzke, R., Xie, Z., Llanos, F., & Chandrasekaran, B. (2018). Tracing the Trajectory of Sensory Plasticity across Different Stages of Speech Learning in Adulthood. *Current Biology*, *28*(9), 1419–1427.e4. https://doi.org/10.1016/j.cub.2018.03.026

Riecke L, Peters JC, Valente G, Kemper VG, Formisano E, Sorger B (2017) Frequency-selective attention in auditory scenes recruits frequency representations throughout human superior temporal cortex. Cereb Cortex 27:3002–3014. 10.1093/cercor/bhw160

Riecke, L., Peters, J. C., Valente, G., Kemper, V. G., Formisano, E., & Sorger, B. (2017). Frequency-Selective Attention in Auditory Scenes Recruits Frequency Representations Throughout Human Superior Temporal Cortex. *Cerebral Cortex*. Retrieved from https://doi.org/10.1093/cercor/bhw160

Roark, C. L., Paulon, G., Rebaudo, G., McHaney, J. R., Sarkar, A., & Chandrasekaran, B. (2024). Individual differences in working memory impact the trajectory of non-native speech category learning. *PloS One*, *19*(6), e0297917. https://doi.org/10.1371/journal.pone.0297917

Roy, J. E., Riesenhuber, M., Poggio, T., & Miller, E. K. (2010). Prefrontal cortex activity during flexible categorization. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *30*(25), 8519–8528. https://doi.org/10.1523/JNEUROSCI.4837-09.2010

Saenz, M., Buracas, G. T., & Boynton, G. M. (2002). Global effects of feature-based attention in human visual cortex. *Nature neuroscience*, *5*(7), 631–632. https://doi.org/10.1038/nn876

Serences, J. T., & Boynton, G. M. (2007). The representation of behavioral choice for motion in human visual cortex. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *27*(47), 12893–12899. https://doi.org/10.1523/JNEUROSCI.4021-07.2007

Serre, T., Wolf, L., Bileschi, S., Riesenhuber, M., & Poggio, T. (2007). Robust object recognition with cortex-like mechanisms. *IEEE transactions on pattern analysis and machine intelligence*, *29*(3), 411–426. https://doi.org/10.1109/TPAMI.2007.56

Sigala, N., & Logothetis, N. K. (2002). Visual categorization shapes feature selectivity in the primate temporal cortex. *Nature*, *415*(6869), 318–320. https://doi.org/10.1038/415318a

Sigala, N., & Logothetis, N. K. (2002). Visual categorization shapes feature selectivity in the primate temporal cortex. *Nature*, *415*(6869), 318–320. https://doi.org/10.1038/415318

Todorov, A., Pakrashi, M., & Oosterhof, N. (2009). Evaluating faces on trustworthiness after minimal time exposure. *Social Cognition*, 27, 813–833. http://doi.org/10.1521/soco.2009.27.6.813

Treue, S., & Maunsell, J. H. (1999). Effects of attention on the processing of motion in macaque middle temporal and medial superior temporal visual cortical areas. The Journal of neuroscience : the official journal of the Society for Neuroscience, 19(17), 7591–7602. https://doi.org/10.1523/JNEUROSCI.19-17-07591.1999

van der Linden, M., van Ruennout, M., & Idefrey, P. (2010). Formation of category representations in superior temporal sulcus. Journal of Cognitive Neuroscience, 22, 1270–1282.

Van Gulick, A. E., & Gauthier, I. (2014). The perceptual effects of learning object categories that predict perceptual goals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(5), 1307–1320. https://doi.org/10.1037/a0036822

Yin, P., Strait, D. L., Radtke-Schuller, S., Fritz, J. B., & Shamma, S. A. (2020). Dynamics and Hierarchical Encoding of Non-compact Acoustic Categories in Auditory and Frontal Cortex. *Current biology : CB*, *30*(9), 1649–1663.e5. https://doi.org/10.1016/j.cub.2020.02.047

Yin, P., Strait, D. L., Radtke-Schuller, S., Fritz, J. B., & Shamma, S. A. (2020). Dynamics and Hierarchical Encoding of Non-compact Acoustic Categories in Auditory and Frontal Cortex. *Current biology : CB*, *30*(9), 1649–1663.e5. https://doi.org/10.1016/j.cub.2020.02.047

Yoo, S. A., Martinez-Trujillo, J. C., Treue, S., Tsotsos, J. K., & Fallah, M. (2022). Attention to visual motion suppresses neuronal and behavioral sensitivity in nearby feature space. *BMC biology*, *20*(1), 220. https://doi.org/10.1186/s12915-022-01428-7

Young, A. W., Frühholz, S., & Schweinberger, S. R. (2020). Face and Voice Perception: Understanding Commonalities and Differences. *Trends in Cognitive Sciences*, *24*(5), 398–410. https://doi.org/10.1016/j.tics.2020.02.001

Yurovsky, D., & Frank, M. C. (2017). Beyond naïve cue combination: salience and social cues in early word learning. *Developmental Science*, *20*(2), 10.1111/desc.12349. https://doi.org/10.1111/desc.12349

Yurovsky, D., Boyer, T. W., Smith, L. B., & Yu, C. (2013). Probabilistic cue combination: less is more. *Developmental Science*, *16*(2), 149–158. https://doi.org/10.1111/desc.12011

Zhong, L., Zhang, Y., Duan, C. A., Deng, J., Pan, J., & Xu, N. L. (2019). Causal contributions of parietal cortex to perceptual decision-making during stimulus categorization. *Nature neuroscience*, *22*(6), 963–973. https://doi.org/10.1038/s41593-019-0383-6

Acknowledgments

The work was supported by grants from National Institutes of Health (R01DC004674) to LLH AT, and FD, the National Science Foundation (SBE/BCS 2414066, 2420979) to LLH and FD, an individual NRSA from the National Institute on Deafness and Communication Disorders to SL (F32DC020625) under the mentorship of LLH and FD. The authors thank Chisom Obasih for her contributions in creating the novel acoustic soundscape.

Competing Interests

The authors declare that they have no competing interests.

Author Contributions

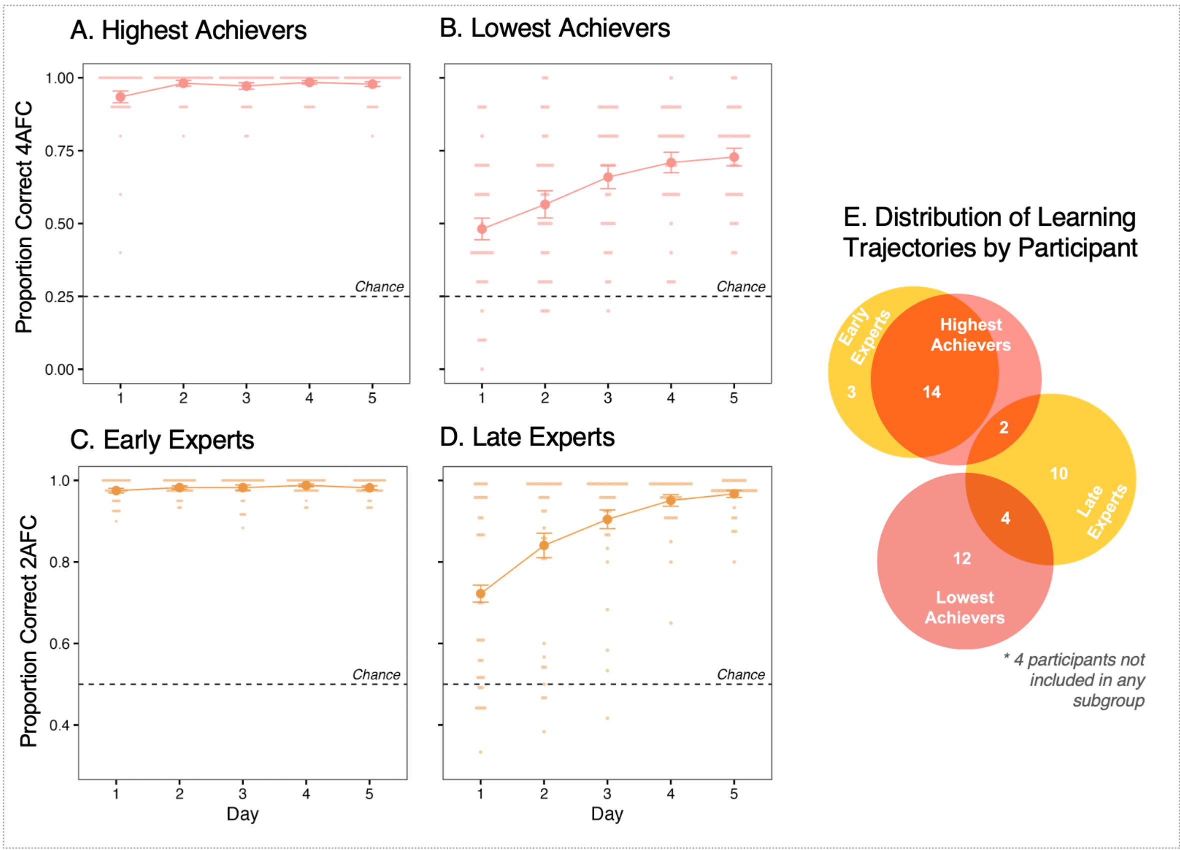
SL contributed to Conceptualization, Methodology, Software, Formal Analysis, Investigation, Data Curation, Writing – Original Draft, Writing – Review and Editing, Visualization. **RR** contributed to Investigation, Data Curation, Writing – Original Draft. **ATT** contributed to Conceptualization, Methodology, Investigation, Funding Acquisition. **LLH** contributed to Conceptualization, Methodology, Investigation, Writing – Original Draft, Writing – Review and Editing, Visualization, Supervision, Resources, Project Administration, Funding

Acquisition. **FD** contributed to Conceptualization, Methodology, Formal Analysis, Investigation, Writing – Original Draft, Writing – Review and Editing, Visualization, Supervision, Project Administration, Resources, Funding Acquisition.

Materials and Correspondence

Correspondence and materials requests should be directed to Fred Dick (f.dick@ucl.ac.uk).

Supplemental Materials



Supplemental Figure S1. There were multiple trajectories of learning to achieve expertise in alien identity categorization. A. Highest Achievers. Defined as participants in the top tercile of performance on the 4AFC generalization task (N=16). **B. Lowest Achievers.** Defined as participants in the bottom tercile of performance on the 4AFC generalization task (N=16). **C. Early Experts.** Defined as participants who, on Day 1, achieved at least 90% accuracy on 2AFC for each category pair defined by diagnostic frequency band (N=17). **D. Late Experts.** Defined as participants who improved 20-55% in accuracy on 2AFC for the low-frequency diagnostic band over Days 1-5 (N=16). **E. Distribution of Learning Trajectories by Participant.** The Venn diagrams show how participants relate to the four learning trajectories. Note that 4 of the 49 total participants are not included in any subgroup based on these criteria.