SUPPLEMENTARY FIGURE 1

A small minority of HVC\textsubscript{X} cells expressed responses to changes in note duration that were not categorical in nature. In 3 of 22 HVC\textsubscript{X} cells tested (2 birds), the auditory response did not fulfill the criteria to establish a response as categorical (see Methods). These data reveal no evidence of neurons selectively tuned to intermediate note durations, indicating that the population of HVC\textsubscript{X} neurons does not represent note duration continuously, further supporting our claim that HVC\textsubscript{X} neurons encode note duration in categorical manner.
SUPPLEMENTARY FIGURE 2

A linear model of auditory responses of HVC\(_x\) neurons to notes of different duration does not account for the observed dataset. We compared the relations between note duration and strength of response for HVC\(_x\) cells in our dataset against the null hypothesis that those cells represented note duration in a linear manner (i.e., HVC\(_x\) cells expressed a linear relation rather than the step-like transition characteristic of categorical activity). In this noisy linear model, the neural response is expected to be minimal at the shortest note duration we tested (normalized response = 0 at note duration 4 ms) and is expected to be maximal at the longest note duration we tested (normalized response = 1 at note duration 31 ms; stated in equation form: \(Y = mX + b\), where \(m = 0.037\) and \(b = -0.148\)). Random variance is included in the model to simulate local transitions with slopes much steeper than the mean value or possibly negative slopes (see Supplementary Methods). Because our dataset included both positive and negative slopes at the categorical boundary (e.g., Fig. 3D, Supp. Fig. 3), we used the absolute value of the slope in the tuning curve of each HVC\(_x\) neuron to compare the steepest slope of each transition in our dataset against the results of the model (shown in panel a). Artificial data created using this model were compared against the dataset of 10 HVC\(_x\) neurons for which we had a high-resolution estimate of the categorical boundary (Supp. Fig. 3). Three epochs of note duration were considered: 1) within the putative short-note category (X <= 14 ms), 2) within the putative long-note category (X >= 27 ms), and 3) values around the putative boundary (14 < X < 27 ms). (a) Within the putative boundary region, the slopes of the steepest transition for cells in our dataset (filled bar, mean ± SE) were significantly steeper than the slope of the model (open bar, \(p = 0.002\), paired t-tests). (b) Within each category, the slopes of responses in our dataset (filled bar) were significantly different than the slope of the model (open bar, \(p = 0.02\)) and were indistinguishable from a slope of zero \((p = 0.94\), no absolute values used in within-category comparison of model and actual values). Together, these data indicate that a linear model cannot account for our observations and that,
consistent with a categorical representation, the population activity of HVC_\chi neurons in response to changes in note duration expresses a steep transition between two regions with slopes indistinguishable from zero.
SUPPLEMENTARY FIGURE 3

The categorical boundary estimated from a subset of HVC<sub>X</sub> neurons tested using high-resolution stimuli corroborates the boundary estimated using the full dataset. A subset of 10 HVC<sub>X</sub> neurons (2 birds, all 10 cells shown here) was tested to probe the categorical boundary at a high resolution, revealing an estimated categorical boundary (filled triangle, 20 ± 4 ms) very similar to that estimated using the full dataset (21 ± 4 ms; p = 0.47, unpaired t-test). Categorical responses were observed in HVC<sub>X</sub> cells regardless of whether the duration of note that was replaced was naturally short (thick lines) or naturally long (thin lines).
SUPPLEMENTARY FIGURE 4

Behavioral testing of one swamp sparrow for which neural data were also collected revealed a correspondence between the perceptual and neurophysiological boundaries. Song stimuli were presented as described in the Methods for experiments performed in the field, and the bird’s behavior was observed using a small camera placed inside the neurophysiological recording chamber. The bird was conditioned to song stimuli using the song of another swamp sparrow, and the number of aggressive displays (crest raises\textsuperscript{46} or wing waves\textsuperscript{7} quantified as described in the Methods) was strongly habituated at time 0. Changing the stimulus to the bird’s own song (12 minutes) evoked a dishabituation of response, indicating that the bird perceived the new song as different than the song to which it had been conditioned. In subsequent song changes in this habituation/dishabituation paradigm (see Methods), this bird responded as if he perceived differences between stimuli that crossed the neurophysiological boundary (i.e., changing note duration from 7 to 31 ms at 98 minutes and from 27 to 5 ms at 218 minutes) but not between stimuli that did not cross the boundary (i.e., changing from 31 to 27 ms at 192 minutes). The bird from which we obtained these behavioral results is the same bird from which we obtained the neural in Figs 2B–C (left and middle columns), Fig 3A–B and Supp. Fig. 5.
SUPPLEMENTARY FIGURE 5

Categorical auditory responses were observed in HVC<sub>X</sub> cells that also expressed a precise sensorimotor correspondence<sup>20</sup>. (a) The auditory response of this HVC<sub>X</sub> neuron was selective for the primary song type (N = 5 cells, 2 birds; data from the same cell shown in all panels; top: auditory response PSTH, 10 ms binsize; bottom: stimulus syllable spectrogram; left: primary song type; center: reverse playback of the primary song type; right: another song type in the bird’s vocal repertoire). (b) Manipulation of the duration of note C in each syllable of the primary song type resulted in a categorical auditory response (data as in Fig. 3A). (c) This cell also expressed a precise correspondence in the timing of action potentials generated when the bird sang the primary song type (top) and during the auditory response (middle) to the primary song type (spectrogram, bottom). The presence of a neural correlate of the animal’s perception in auditory-vocal “mirror neurons<sup>20b</sup>” strengthens the idea that cells expressing an auditory-vocal correspondence may facilitate perception of the signals used in vocal communication.
SUPPLEMENTARY FIGURE 6

Responses of individual birds in behavioral testing of the perceptual boundary. Swamp sparrows in the Pennsylvania population perceived strong differences when the transition in note duration spanned the neurophysiological boundary detected in HVC<sub>x</sub> neurons of Pennsylvania birds (right, N = 8 birds, statistics as reported in legend of Fig. 4C) but perceived little or no difference when the transition spanned the perceptual boundary detected in New York birds (middle, N = 8 birds) or spanned no putative boundary (control, left, N = 8 birds). The y-axis indicates the difference between the number of territorial displays evoked in the first block of testing using the dishabituation test stimulus and the number of displays evoked in the final block of testing using the habituation stimulus (thus negative values are possible, see Methods for details). Small values are consistent with the bird perceiving little or no difference between the habituation and dishabituation stimuli, and large positive values are consistent with the bird perceiving the two stimuli as different. These data are summarized in Fig. 4C.
SUPPLEMENTARY METHODS

Animal Collection and Care

Methods for the collection of swamp sparrows from the wild and their care in captivity through compliance with recommendations of Duke University Animal Care and Use Committee and state and federal regulations governing the capture and use of wild birds have been described in detail previously\textsuperscript{20}. All males used to measure the categorical boundary of the neural response were collected as adults (age > 1 year) from their breeding grounds in Crawford County, PA.

Song Stimulus Preparation

Gradually lengthening photoperiod (1 hr/week from 9:15 up to 15:9 L:D cycle) meant to simulate the onset of the breeding season, together with subcutaneous implantation of testosterone, were used to induce the birds to sing\textsuperscript{20}. Each bird’s full repertoire of songs (2 to 5 song types) was recorded\textsuperscript{20}, and the fully crystallized state of these songs served as further evidence that the birds were fully adult. Exemplars of each song type were digitized\textsuperscript{20} to be used as stimulus songs and to be used in the construction of stimulus songs to test categorical responsiveness. The characteristic syllable of each song type in the bird’s repertoire was examined for notes of the type that are known from behavioral studies\textsuperscript{7} to be perceived categorically (Category I vs. Category VI in the nomenclature of Marler and Pickert\textsuperscript{12}). Notes of these categories served as “target notes” (e.g., note C in Fig. 3A) for replacement using notes from songs of other swamp sparrows, and a target note could occur at any place in the syllable. Each replacement note had similar spectral characteristics but different duration than the target note (e.g., Fig. 3A), and only one target note in the syllable was replaced in any stimulus. The possible durations of replacement notes were 4, 5, 7, 8, 16, 19, 22, 25, 27, and 31 ms. These values were chosen to replicate as closely as possible the methods used in the previous behavioral assessment of categorical perception in swamp sparrows\textsuperscript{7} and to resolve the location of the categorical boundary. The resulting synthetic syllable...
contained the replacement note inserted in place of the target note, and all other notes and internote intervals were identical to the natural song type. This synthetic syllable was then assembled into a trill to form a stimulus song with normal duration and trill rate identical to the natural song (intersyllable interval was adjusted as required to match the natural trill rate). Thus, replacement of one target note in a song type resulted in a set of up to 11 variants of that song type. This procedure was repeated for each candidate target note in each song type in the bird’s repertoire, resulting in the preparation of a large set of potential stimuli for each bird.

We used replacement note durations that differed in approximately logarithmic increments (e.g., Fig. 3A) because perceived changes in acoustic stimuli, including temporal changes, are best described as following a logarithmic scale according to the Weber-Fechner law (i.e., equal intervals on a log scale are equally discriminable), and logarithmic differences have been used to test temporal discrimination and categorical perception\textsuperscript{7,47-50}. An additional advantage is that logarithmic increments were used in the field study that established categorical perception in swamp sparrows\textsuperscript{7}, thus enabling direct comparison of our findings with previous behavioral work.

In a subset of 10 HVC\textsubscript{X} cells, additional synthetic stimuli were presented in which either the rate of frequency modulation (FM) or the frequency bandwidth (BW) was controlled while the duration of the note was varied (Figs. 5A-C). After a cell had been determined to express a categorical response (see below), these additional synthetic stimuli were presented to determine whether FM or BW played an important role in directing the categorical nature of the response, or whether note duration was the salient feature of the stimulus (resulting in a stimulus set of up to 16 variants of the primary song type for these 10 cells).
Microdrive Implantation Surgery

All experiments were performed in a sound attenuating chamber in which the bird was housed both during testing and overnight. Neurons were sampled using a miniaturized micromanipulation device in awake and freely behaving birds. Several days prior to implantation, birds were transferred from their housing cage to the recording chamber, a sound-attenuating box (Acoustic Systems) where they would reside throughout experimentation. During implantation, adult male swamp sparrows were anesthetized using isoflurane (inhalation, 1-3% in 100% O₂) and placed in a stereotaxic device. A small incision was made in the skin overlying the skull, and the outer leaflet of bone was removed over HVC, Area X and RA. A small craniotomy (approximately 300 by 300 microns) was made in the inner leaflet over Area X, and a small custom-made bipolar stimulus electrode (JF Prather) was inserted to the proper depth. The implant site was covered with a sterile film and the electrode was secured using dental cement. With the electrode in Area X firmly secured, the head was repositioned and the same implant procedure was repeated to place a bipolar stimulus electrode in RA. With both stimulus electrodes firmly in place, another small craniotomy was made directly over HVC. HVC was located by passing brief (~ 100 μsec) current pulses through the stimulating the electrodes in Area X and RA and using a sterilized extracellular electrode (Carbostar 1, Kation Scientific) to observe the extent of the region expressing the resultant antidromic “hash.” The microdrive recording device was implanted so that the recording electrodes were initially positioned slightly dorsal of HVC. The microdrive was secured to the skull using dental cement (microdrive ~ 1.2 g including dental cement, birds ~ 16 g), and the incision site was closed using surgical skin adhesive (Vetbond). The bird was monitored closely until it was fully recovered, typically < 15 min. After the recording session was complete (1-5 weeks), the bird was deeply anesthetized with equithesin, perfused transcardially with saline and then 4% paraformaldehyde, and the brain was processed histologically. All electrode positions were verified at the end of each experiment using Nissl-stained sagittal sections (thickness 75 microns).
**Experimental Protocol**

Birds were allowed to recover for three days following the implantation procedure before recording began. During electrophysiological recording, microdrive electrodes were slowly advanced into HVC while weak electrical stimulation was delivered to the stimulus electrodes in either Area X or RA (100 μsec pulses, ~ 100 μA). The boundaries of HVC could be reliably identified by observing where antidromic activity was evident. Once an electrode was positioned in HVC, the electrode was advanced very slowly so that antidromically-evoked action potentials of individual neurons could be identified. All neural data were amplified, filtered (band pass 500 Hz to 10 kHz), and digitized (25 kHz) to computer file (LabView).

Action potentials of individual units were discriminated using amplitude discrimination of the largest unit in a record (custom software) or discrimination based on waveform characteristics (WaveClus). In both cases, single unit isolation was verified using an interspike interval histogram to test for the presence of a refractory period. Individual units were identified using antidromic stimulation via the electrodes placed in Area X and RA or by their characteristic electrophysiological response properties. In antidromic identification, HVC\textsubscript{X} units displayed fixed-latency action potential responses to stimulation in Area X but no response to stimulation in RA. In contrast, HVC\textsubscript{RA} units displayed fixed-latency action potential responses to stimulation in RA but not in Area X. Each of these classes of projection neuron could be distinguished from HVC interneurons, which expressed variable-latency responses to stimulation in either RA or Area X and occasionally to stimulation at both sites.
When a single unit had been isolated and identified, stimulus presentation was immediately initiated (10 sec quiet interval between each song presentation, stimuli presented in randomized order). Songs were played to the sparrow at 70 dB (peak RMS, A-weighted) through a speaker placed 20 to 35 cm away in the chamber (distance varied according to the bird’s location in the cage). Playback of the bird’s entire song repertoire, as well as their synthetic variants (see below), were used to assess the auditory response of each neuron described in the Results. We enforced the following criteria to qualify a neuron as suitable for further analysis: 1) action potentials must have been reliably distinguishable as belonging to only a single unit, and 2) all song types in the bird’s repertoire must have been presented as auditory stimuli.

Song stimuli consisted of natural song types and synthetic variants of those types from the experimental subject. Natural song types (unaltered from the original recordings) were used to assess the auditory selectivity of each neuron (i.e., identify the “primary song type”). With the primary song type identified for a given cell, playback of the natural song type and synthetic variants (see Song Stimulus Preparation, above) was immediately initiated (10 sec quiet interval between each song presentation, stimuli presented in randomized order). Songs were played to the sparrow at 70 dB (peak RMS, A-weighted; speaker placed 20 – 35 cm away in the chamber, distance varied according to the bird’s location in the cage).

We enforced the following criteria to establish a neuron as suitable for analysis: 1) action potentials must have been reliably distinguishable as belonging to only a single unit, 2) in tests of categorical response, the response to the natural song type must be present both before and after tests using synthetic stimuli; this ensured that the same cell was maintained throughout testing and served as a positive control for cases of synthetic stimuli that evoked no auditory response, and 3) each auditory stimulus must have been tested by presentation of at least 40 syllables; although nearly all
cells greatly exceeded this criterion (e.g., Fig. 2C), it nonetheless ensured that assessment of auditory responsiveness was based on a sufficient dataset. Extracellular recordings were collected from 29 individual HVC<sub>x</sub> units (5 birds) identified using either antidromic stimulation (significant auditory responses in 10 of 12 antidromically identified HVC<sub>x</sub> cells, 2 birds) or auditory response characteristics (significant auditory responses in 14 of 17 cells, 3 birds). Extracellular recordings were also collected from 18 individual HVC<sub>int</sub> units (the same 5 birds as those from which HVC<sub>x</sub> cells were sampled; cells identified using either antidromic stimulation (N = 5 cells, 2 birds) or characteristics of the action potential shape and auditory response (N = 13 cells, 3 birds) that met these criteria. Specifically, HVC<sub>int</sub> cells were very active even in the absence of a stimulus, they had broadly tuned responses across song types, they expressed tonic firing in response to stimulus presentation, and the shape of the trajectory of their extracellularly recorded afterhyperpolarization was concave in appearance. In contrast, HVC<sub>x</sub> neurons were much less active in the absence of a stimulus, had highly selective responses across song types, expressed phasic firing in response to stimulus presentation, and the shape of their afterhyperpolarization was convex in appearance (Prather et al., in preparation).

**Data Analysis**

All analyses were performed in Matlab using custom software (JF Prather and S Nenkov). Construction of raster and histogram plots aligned to either the whole song (e.g., Fig. 2B) or each syllable in the song (e.g., Fig. 2C) has been described previously<sup>20</sup>.

**Quantification of Auditory Response**

Auditory responses of HVC<sub>x</sub> neurons to presentation of the primary song type were quantified as action potentials per bin in the peri-stimulus time histogram (PSTH, e.g., Fig. 2C, left) and compared against the background firing rate of the same cell when no stimulus was present. The
mean background rate plus 5 SD was taken as the threshold for significance. If the value in any bin in the PSTH exceeded that threshold (accounting for bin size), the auditory response of that HVC$_x$ cell was deemed significant. This threshold-crossing means of assessing the significance of auditory responses is an effective means of describing neurons with the phasic, temporally precise activity characteristic of HVC$_x$ cells (e.g., Fig. 2C, left). However, the responses of HVC interneurons typically consisted of tonic increases in firing rate throughout the stimulus presentation (Prather et al., in preparation) and were therefore tested for significance using response strength, a metric that compares the activity evoked by a stimulus to the activity present in the same cell in the baseline condition$^{37}$. Interneurons were considered valid for categorical testing if they expressed a significant response to the natural song type (quantified as action potentials per syllable, response strength $p < 0.05$). In assessment of auditory responses to stimuli used to assess categorical activity, the responses of HVC$_x$ neurons or HVC$_{int}$ cells were normalized using the strongest response to any member of the set of synthetic stimuli. The resulting normalized values, referred to as the “normalized response” (e.g., Fig. 3B-D), enabled comparison across cells (e.g., Fig. 3B, left) and birds (e.g., Fig. 3C, left), and comparison of the relative tendencies of HVC$_x$ neurons and HVC$_{int}$ cells to express categorical responses (e.g., Figs 3B-C).

**Neural Assessment of Categorical Responsiveness**

In preliminary analysis, categorical responsiveness was tested by comparing responses to stimuli containing replacement notes with duration less than the previously reported perceptual boundary ($\sim 13$ ms$^3$) against responses to stimuli containing replacement notes with duration greater than that boundary. However, we noted that responses to stimuli with a replacement note duration of 16 ms were consistently like those evoked by much shorter notes (e.g., Fig. 3A), suggesting that the previously reported categorical boundary ($\sim 13$ ms) might be invalid for our dataset. Therefore we sought a means of assessing categorical responsiveness independent of that previous measure.
Noting that the dataset of auditory responses in HVCx cells tended to include transitions from strong responses (normalized responses closer to 1) to little or no responses (normalized responses closer to 0), we used interpolation to compute the note duration at which the auditory response crossed 0.5. In cases wherein interpolated data crossed 0.5 multiple times (e.g., green line in Fig. 3D), the transition point in the cell was represented by the mean of the note durations corresponding to those multiple crossing points. This method was used to compute a putative boundary in the auditory responsiveness of these cells (~ 21 ms). This value was then used to consider whether cells expressed categorical responses to stimuli on either side of this putative boundary, and this knowledge enabled us to obviate possible confounds due to inclusion of cases with replacement note durations near this putative boundary (i.e., note durations from 14 to 27 ms).

The following criteria were enforced in off-line analysis to determine whether a given cell expressed a categorical auditory response: 1) auditory response of the cell to synthetic variants of the primary song type must have been tested using 2 or more stimuli with replacement notes in the “short note” category (<= 14 ms) and 2 or more stimuli in the “long note” category (>= 27 ms); this ensured that response to each group of stimuli was adequately tested while also avoiding possible complications of including stimuli with replacement note durations close to the putative categorical boundary, 2) because a normalized response strength of 0.5 was taken as the transition between the presence and the absence of a strong response (see above), cells were said to express categorical responses if the normalized responses to all members of one group of stimuli (e.g., all cases with replacement notes <= 14) were all on one side of the transition (e.g., all responses greater than 0.5) while all members of the remaining group (e.g., all cases with replacement notes >= 27 ms) were all on the other side of the transition (e.g., all responses less than 0.5). Cells that expressed categorical responses as defined using these criteria are shown in Figs. 3A-D. Cells that did not meet these criteria were said to express auditory responses that were not categorical (e.g.,
not all responses to all notes <= 14 ms greater than 0.5, data not shown). In short, responses were said to be categorical if activity was similar among members of a group of stimuli, but different between groups of stimuli, indicating a greater sensitivity to stimulus category than to the physical properties of the stimulus.

These criteria are in close agreement with previous standards for what constitutes a categorical response, which include: “auditory contrast between two notes on opposite sides of the… boundary should be more salient than equally different pairs of within-category notes taken from either side of the boundary”, and regarding the responses of stimulus-selective neurons to different test stimuli, there should be a “difference in activity between the categories, but… similar… within each category, indicating greater sensitivity to stimulus category than to identity”. Furthermore, our criteria are in line with standards used in studies of human categorical perception of speech. Those criteria were: 1) distinct labeling categories with sharp boundaries, 2) no discrimination between stimuli from the same category, 3) a peak in discrimination at category boundaries, and 4) close agreement between actual discrimination and that predicted from the labeling results assuming absolute categorization. Our data address the criteria of Studdert-Kennedy et al. in the following ways: 1) distinct perceptual categories are shown in Fig. 4C and have been shown previously by Nelson and Marler. In addition, a sharp boundary in the neural responses is shown in Fig. 3D and is quantified (21 ± 4 ms) in the text, 2) no discrimination between stimuli from the same category is evident in Fig. 4C in the similarity of the responses within-category for PA birds (stimulus set 2) versus control (stimulus set 1). A similar trend is evident in the neural data shown in Figs. 3A-C, 3) a peak in discrimination at category boundaries is shown in Fig. 4C in the perception of differences when the neural boundary was crossed (stimulus set 3) but not when the neural boundary was not crossed (stimulus sets 1 and 2), 4) while the inability of birds to report their perception as fully as human subjects gave us no
“labeling results” per se, the neural data served a relevant role by providing a prediction of what should have been observed in actual discrimination, and there was close agreement between actual discrimination and that predicted by the neural data. Both our neural and behavioral data (see below) meet each of these previous standards for what constitutes a categorical response, thus establishing our results as evidence of categorical neural activity and categorical perception.

We compared our claims of categorical transitions in the tuning curves of individual HVC$_x$ neurons (e.g., Fig. 3D, Supp. Fig. 3) against the null hypothesis that those cells represented note duration in a linear manner (i.e., HVC$_x$ cells expressed a linear relation rather than the step-like transition characteristic of categorical activity). In this noisy linear model (Supp. Fig. 2), the neural response is expected to be minimal at the shortest note duration we tested (normalized response = 0 at note duration 4 ms) and is expected to be maximal at the longest note duration we tested (normalized response = 1 at note duration 31 ms; stated in equation form: $Y = mX + b$, where $m = 0.037$ and $b = -0.148$). Three epochs of note duration were considered: 1) within the putative short-note category ($X <= 14$ ms), 2) within the putative long-note category ($X >= 27$ ms), and 3) values around the putative boundary ($14 < X < 27$ ms). In the putative boundary region, the slopes of the steepest transition for cells in our dataset included both positive and negative slopes at the categorical boundary (e.g., Fig. 3D, Supp. Fig. 3), therefore we used the absolute value of slopes in the tuning curve of each HVC$_x$ neuron to compare the steepest slope of each transition in our dataset against the results of the model (Supp. Fig. 2A). Random variance was also included in the model to simulate local transitions with slopes much steeper than the mean value (model slopes modeled with Poisson distribution; values generated using linear mean (0.037) multiplied by values from a Poisson distribution with mean = 1; resulting model slope mean ± SE in Supp. Fig. 2A: 0.043 ± 0.009, range: 0 to 0.148). Within the putative short-note and long-note categories, both
positive and negative slopes could be considered for all cells, therefore no absolute value was applied to the values in our dataset. Model data for these comparisons were simulated using a Normal distribution, which was necessary to allow for the possibility of negative slopes (values generated from a Normal distribution with mean = 0.037 and standard deviation = 0.037; resulting model slope mean ± SE in Supp. Fig. 2B: 0.040 ± 0.008, range: -0.031 to 0.090). Together, these comparisons allowed us to test whether a linear model could account for our observations, or whether, consistent with a categorical representation, the population activity of HVCx neurons in response to changes in note duration expresses a steep transition between two regions with slopes indistinguishable from zero (Supp. Fig. 2).

**Behavioral Assessment of Categorical Perception**

All field tests were performed in Crawford County, PA during May and June of 2007. The Pennsylvania State Game Commission gave permission for access to field sites, and the Pymatuning Laboratory of Ecology, University of Pittsburgh, gave logistical support. Methods closely paralleled the previous study of categorical perception in a New York population of swamp sparrows7. Stimulus songs were made from natural songs recorded in previous years from swamp sparrows from the same population, and the set of stimuli comprised synthetic variants of 8 songs recorded from 6 birds. A target note in each of the 8 songs was replaced by one of 4 natural notes from another swamp sparrow’s song (4, 8, 16, 32 ms). Synthetic songs were assembled as described above, and those stimuli were arranged to be presented in pairs (4:8 ms, 8:16 ms, 16:32 ms), with the sequence of presentation for each pair randomized. One pair of songs was presented to each bird through a speaker placed inside of the bird’s breeding territory, and perception of song features was assessed using a habituation/dishabituation paradigm (see below). To avoid the possible confound of stimuli presented on one day accidentally habituating the response of a neighboring bird, care was taken to ensure that the same song was not presented to birds within ~
200 m of each other unless at least 2 days separated the trials. The stimuli that we presented (variants containing notes of 4, 8, 16, 32 ms) allowed us to test the Pennsylvania birds’ perception of stimuli that crossed the boundary detected in neural recordings of Pennsylvania birds but did not cross the boundary reported previously for New York birds (16:32 ms, N = 8 birds), stimuli that crossed the perceptual boundary reported for New York birds but did not cross the boundary detected in our neural recordings (8:16, N = 8 birds), and stimuli that crossed no known perceptual boundary (4:8 ms, N = 8 birds).

Duplicating the experimental methods employed previously\(^7\), a single stimulus song was presented at the rate of 4 songs per minute over the course of a 3-minute block, followed by a 3-minute interval of silence. Territorial behaviors (obvious displays characterized by waving of the wings) during the stimulus block were observed by one experimenter while another experimenter recorded the time at which each behavior occurred. The 3-minute block was subdivided into 5-second blocks and each subunit was scored with a 1 if any behavioral expression occurred during that time and a 0 otherwise. The score was computed (maximum score: 36) for the first stimulus block in which a given song was presented, and the same stimulus song was presented in sequential stimulus blocks until the subject bird’s response had habituated to 25\% of the original response and remained below that value for two consecutive stimulus blocks. At that point, the stimulus song was switched to the other member of the stimulus pair in the following stimulus block. In the final 3-minute block following the switch, the number of 5-second subunits in which territorial behavior was expressed were tallied and compared against the number expressed in the final block of habituation, and this value was taken as an index of perceived difference (y-axis in Fig. 4C). In this habituation/dishabituation paradigm, perception of a difference between the two songs typically resulted in a return of vigorous expression of territorial behaviors, whereas no perception of a difference between the stimuli resulted in persistent habituation of response.


