

Emergence of invariant representation of vocalizations in the auditory cortex

Isaac M. Carruthers,^{1,2} Diego A. Laplagne,³ Andrew Jaegle,^{1,4} John J. Bruguglio,^{1,2}
Laetitia Mwilambwe-Tshilobo,¹ Ryan G. Natan,^{1,3} and Maria N. Geffen^{1,2,4}

¹Department of Otorhinolaryngology and Head and Neck Surgery, University of Pennsylvania, Philadelphia, Pennsylvania; ²Graduate Group in Physics, University of Pennsylvania, Philadelphia, Pennsylvania; ³Brain Institute, Federal University of Rio Grande do Norte, Natal, Brazil; and ⁴Graduate Group in Neuroscience, University of Pennsylvania, Philadelphia, Pennsylvania

Submitted 28 January 2015; accepted in final form 25 August 2015

Carruthers IM, Laplagne DA, Jaegle A, Bruguglio JJ, Mwilambwe-Tshilobo L, Natan RG, Geffen MN. Emergence of invariant representation of vocalizations in the auditory cortex. *J Neurophysiol* 114: 2726–2740, 2015. First published August 26, 2015; doi:10.1152/jn.00095.2015.—An essential task of the auditory system is to discriminate between different communication signals, such as vocalizations. In everyday acoustic environments, the auditory system needs to be capable of performing the discrimination under different acoustic distortions of vocalizations. To achieve this, the auditory system is thought to build a representation of vocalizations that is invariant to their basic acoustic transformations. The mechanism by which neuronal populations create such an invariant representation within the auditory cortex is only beginning to be understood. We recorded the responses of populations of neurons in the primary and nonprimary auditory cortex of rats to original and acoustically distorted vocalizations. We found that populations of neurons in the nonprimary auditory cortex exhibited greater invariance in encoding vocalizations over acoustic transformations than neuronal populations in the primary auditory cortex. These findings are consistent with the hypothesis that invariant representations are created gradually through hierarchical transformation within the auditory pathway.

auditory cortex; hierarchical coding; invariance; processing; vocalizations

IN EVERYDAY acoustic environments, communication signals are subjected to acoustic transformations. For example, a word may be pronounced slowly or quickly, or by different speakers. These transformations can include shifts in spectral content, variations in frequency modulation, and temporal distortions. Yet the auditory system needs to preserve the ability to distinguish between different words or vocalizations under many acoustic transformations, forming an “invariant” or “tolerant” representation (Sharpee et al. 2011). Presently, little is understood about how the auditory system creates a representation of communication signals that is invariant to acoustic distortions.

It has been proposed that within the auditory processing pathway, invariance emerges in a hierarchical fashion, with higher auditory areas exhibiting progressively more tolerant representations of complex sounds. The auditory cortex (AC) is an essential brain area for encoding behaviorally important acoustic signals (Aizenberg and Geffen 2013; Engineer et al. 2008; Fritz et al. 2010; Galindo-Leon et al. 2009; Recanzone and Cohen 2010; Schnupp et al. 2006; Wang et al. 1995). Up

to and within the primary auditory cortex (A1), the representations of auditory stimuli are hypothesized to support an increase in invariance. Whereas neurons in input layers of A1 preferentially respond to specific features of acoustic stimuli, neurons in the output layers become more selective to combinations of stimulus features (Atencio et al. 2009; Sharpee et al. 2011). In the visual pathway, recent studies suggest a similar organizing principle (DiCarlo and Cox 2007), such that populations of neurons in higher visual area exhibit greater tolerance to visual stimulus transformations than neurons in the lower visual area (Rust and DiCarlo 2010; Rust and DiCarlo 2012). Here, we tested whether populations of neurons beyond A1, in a nonprimary auditory cortex, support a similar increase in invariant representation.

We focused on the transformation between A1 and one of its downstream targets in the rat, the suprarhinal auditory field (SRAF) (Arnault and Roger 1990; Polley et al. 2007; Profant et al. 2013; Romanski and LeDoux 1993b). A1 receives projections directly from the lemniscal thalamus into the granular layers (Kimura et al. 2003; Polley et al. 2007; Roger and Arnault 1989; Romanski and LeDoux 1993b; Storace et al. 2010; Winer et al. 1999) and sends extensive convergent projections to SRAF (Covic and Sherman 2011; Winer and Schreiner 2010). Neurons in A1 exhibit short-latency, short time-to-peak responses to tones (Polley et al. 2007; Profant et al. 2013; Rutkowski et al. 2003; Sally and Kelly 1988). By contrast, neurons in SRAF exhibit delayed response latencies, longer time to peak in response to tones, spectrally broader receptive fields and lower spike rates in responses to noise than neurons in A1 (Arnault and Roger 1990; LeDoux et al. 1991; Polley et al. 2007; Romanski and LeDoux 1993a), consistent with responses in nonprimary AC in other species (Carrasco and Lomber 2011; Kaas and Hackett 1998; Kikuchi et al. 2010; Kusmierek and Rauschecker 2009; Lakatos et al. 2005; Petkov et al. 2006; Rauschecker and Tian 2004; Rauschecker et al. 1995). These properties also suggest an increase in tuning specificity from A1 to SRAF, which is consistent with the hierarchical coding hypothesis.

Rats use ultrasonic vocalizations (USVs) for communication (Knutson et al. 2002; Portfors 2007; Sewell 1970; Takahashi et al. 2010). Like mouse USVs (Galindo-Leon et al. 2009; Liu and Schreiner 2007; Marlin et al. 2015; Portfors 2007), male USVs evoke temporally precise and predictable patterns of activity across A1 (Carruthers et al. 2013), thereby providing us an ideal set of stimuli with which to probe invariance to acoustic transformations in the auditory cortex. The USVs used in this study are part of the more general class of high-

Address for reprint requests and other correspondence: M. N. Geffen, Dept. of Otorhinolaryngology and Head and Neck Surgery, Univ. of Pennsylvania Perelman School of Medicine, 5 Ravdin, 3400 Spruce St., Philadelphia, PA 19104 (e-mail: mgeffen@med.upenn.edu).

frequency USVs, which are produced during positive social, sexual, and emotional situations (Barfield et al. 1979; Bialy et al. 2000; Brudzynski and Pniak 2002; Burgdorf et al. 2000; Burgdorf et al. 2008; Knutson et al. 1998; 2002; McIntosh et al. 1978; Parrott 1976; Sales 1972; Wohn et al. 2008). The specific USVs were recorded during friendly male adolescent play (Carruthers et al. 2013; Sirotin et al. 2014; Wright et al. 2010). Responses of neurons in A1 to USVs can be predicted based on a linear non-linear model that takes as an input two time-varying parameters of the acoustic waveform of USVs: the frequency- and temporal-modulation of the dominant spectral component (Carruthers et al. 2013). Therefore, we used these sound parameters as the basic acoustic dimensions along which the stimuli were distorted.

At the level of neuronal population responses to USVs, response invariance can be characterized by measuring the changes in neurometric discriminability between USVs as a function of the presence of acoustic distortions. Neurometric discriminability is a measure of how well an observer can discriminate between stimuli based on the recorded neuronal signals (Bizley et al. 2009; Gai and Carney 2008; Schneider and Woolley 2010). Because this measure quantifies available information, which is a normalized quantity, it allows us to compare the expected effects across two different neuronal populations in different anatomical areas. If the representation in a brain area is invariant, discriminability between USVs is expected to show little degradation in response to acoustic distortions. On the other hand, if the neuronal representation is based largely on direct encoding of acoustic features, rather than encoding of the vocalization identity, the neurometric discriminability will be degraded with changes in the acoustic features of the USVs.

Here, we recorded the responses of populations of neurons in A1 and SRAF to original and acoustically distorted USVs, and tested how acoustic distortion of USVs affected the ability of neuronal populations to discriminate between different instances of USVs. We found that neuronal populations in SRAF exhibit greater generalization for acoustic distortions of vocalizations than neuronal populations in A1.

METHODS

Animals. All procedures were approved by the Institutional Animal Care and Use Committee of the University of Pennsylvania. Subjects in all experiments were adult male Long Evans rats, 12–16 wk of age. Rats were housed in a temperature- and humidity-controlled vivarium on a reversed 24 h light-dark cycle with food and water provided ad libitum.

Stimuli. The original vocalizations were extracted from a recording of an adult male Long Evans rat interacting with a conspecific male in a custom-built social arena (Fig. 1A). As described previously (Sirotin et al. 2014), the arena is split in half and kept in the dark, such that the two rats can hear and smell each other and their vocalizations can be unambiguously assigned to the emitting subject. In these sessions, rats emitted high rates of calls from the “50 kHz” family and none of the “22 kHz” type, suggesting interactions were positive in nature (Brudzynski 2009). Recordings were made using condenser microphones with nearly flat frequency response from 10 to 150 kHz (CM16/CMPA-5V, Avisoft Bioacoustics) digitized with a data acquisition board at 300 kHz sampling frequency (PCIe-6259 DAQ with BNC-2110 connector, National Instruments).

We selected eight representative USVs with distinct spectrotemporal properties (Figs. 1 and 2) (Carruthers et al. 2013) from the 6,865 ones

emitted by one of the rats. We contrasted mean frequency and frequency bandwidth of the selected calls with that of the whole repertoire from the same rat (Fig. 2B). We calculated vocalization center frequency as the mean of the fundamental frequency and bandwidth as the root mean square of the mean-subtracted fundamental frequency of each USV. We denoised and parametrized USVs following methods published previously by our group (Carruthers et al. 2013). Briefly, we constructed a noiseless version of the vocalizations using an automated procedure. We computed the noiseless signal as a frequency- and amplitude-modulated tone, such that at any time, the frequency, $f(t)$, and amplitude, $a(t)$, of that tone were matched to the peak amplitude and frequency of the recorded USV at all times, using the relation

$$x(t) = a(t) \sin \left(2\pi \int_0^t f(\tau) d\tau \right).$$

We constructed the acoustic distortions of the 8 selected vocalizations along the dimensions that are essential for their encoding in the auditory pathway (Fig. 1B). For each of these 8 original vocalizations we generated eight different transformed versions, amounting to 9 versions (referred to as transformation conditions) of each vocalization. We then generated the stimulus sequences by concatenating the vocalizations, padding them with silence such that they were presented at a rate of 2.5 Hz.

Stimulus transformations. The 8 transformations applied to each vocalization were temporal compression (designated T–, transformed

by scaling the length by a factor of 0.75: $x(t) = a \left(\frac{t}{0.75} \right) \sin \left[2\pi \int_0^{\frac{t}{0.75}} f(0.75\tau) d\tau \right]$, temporal dilation (T+, length \times 1.25: $x(t) = a$

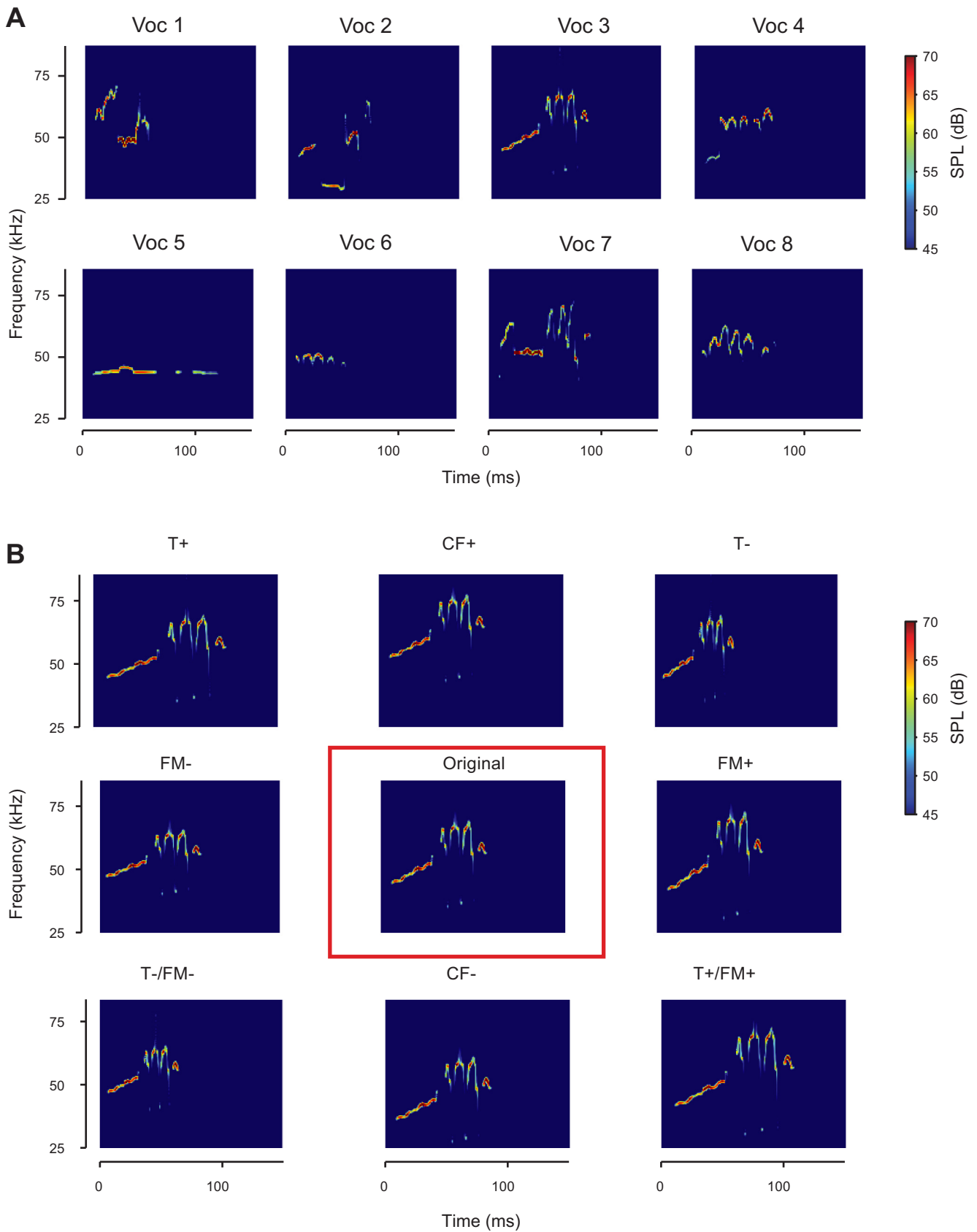
$\left(\frac{t}{1.25} \right) \sin \left[2\pi \int_0^{\frac{t}{1.25}} f(1.25\tau) d\tau \right]$, spectral compression (FM–, bandwidth \times 0.75: $x(t) = a(t) \sin \left(2\pi \int_0^t \{0.75[f(\tau) - \bar{f}] + \bar{f}\} d\tau \right)$, spectral dilation

(FM+, bandwidth \times 1.25: $x(t) = a(t) \sin \left(2\pi \int_0^t \{1.25[f(\tau) - \bar{f}] + \bar{f}\} d\tau \right)$, spectrotemporal compression (T–/FM–, length and bandwidth \times 0.75:

$x(t) = a \left(\frac{t}{0.75} \right) \sin \left(2\pi \int_0^{\frac{t}{0.75}} \{0.75[f(0.75\tau) - \bar{f}] + \bar{f}\} d\tau \right)$, spectrotemporal dilation (T+/FM+, length and bandwidth \times 1.25: $x(t) = a \left(\frac{t}{1.25} \right) \sin$

$\left[2\pi \int_0^{\frac{t}{1.25}} \{1.25[f(1.25\tau) - \bar{f}] + \bar{f}\} d\tau \right)$, center-frequency increase (CF+, frequency + 7.9 kHz: $x(t) = a(t) \sin \left\{ 2\pi \int_0^t [f(\tau) + 7.9 \text{ kHz}] d\tau \right\}$, and center-frequency decrease (CF minus, frequency minus 7.9 kHz: $x(t) = a(t) \sin \left\{ 2\pi \int_0^t [f(\tau) - 7.9 \text{ kHz}] d\tau \right\}$). Spectrograms of denoised vocalizations are shown in Fig. 1A. Spectrograms of transformations of one of the vocalizations are shown in Fig. 1B.

Microdrive implantation. Rats were anesthetized with an intraperitoneal injection of a mixture of ketamine (60 mg/kg body wt) and dexmedetomidine (0.25 mg/kg). Buprenorphine (0.1 mg/kg) was administered as an operative analgesic with ketoprofen (5 mg/kg) as postoperative analgesic. A small craniotomy was performed over A1 or SRAF. Eight independently movable tetrodes housed in a microdrive (6 for recordings and 2 used as a reference) were implanted in A1 (targeting layer 2/3), SRAF (targeting layer 2/3) or both as previously described (Carruthers et al. 2013; Otazu et al. 2009). The microdrive was secured to the skull using dental cement and acrylic. The tetrodes' initial lengths were adjusted to target A1 or SRAF during implantation, and were furthermore advanced by up to 2 mm (in 40- μ m increments, once per recording session) once the tetrode



was implanted. A1 and SRAF were reached by tetrodes implanted at the same angle (vertically) through a single craniotomy window (on the top of the skull) by advancing the tetrodes to different depths on the basis of their stereotactic coordinates (Paxinos and Watson 1986; Polley et al. 2007). At the endpoint of the experiment a small lesion was made at the electrode tip by passing a short current (10 μ A, 10 s) between electrodes within the same tetrode. The brain areas from which the recordings were made were identified through histological reconstruction of the electrode tracks. Limits of brain areas were taken from Paxinos and Watson (1986) and Polley et al. (2007).

Stimulus presentation. The rat was placed on the floor of a custom-built behavioral chamber, housed inside a large double-walled acoustic isolation booth (Industrial Acoustics). The acoustical stimulus was delivered using an electrostatic speaker (MF-1, Tucker-Davis Technologies) positioned directly above the subject. All stimuli were controlled using custom-built software (Mathworks), a high-speed digital-to-analog card (National Instruments) and an amplifier (TDT). The speaker output was calibrated using a 1/4-in. free-field microphone (Bruel and Kjaer, type 4939) at the approximate location of the animal's head. The input to the speaker was compensated to ensure that pure tones between 0.4 and 80 kHz could be output at a volume of 70 dB to within a margin of at most 3 dB. Spectral and temporal distortion products as well as environmental reverberation products were >50 dB below the mean sound pressure level relative to 20 μ Pa (SPL) of all stimuli, including USVs (Carruthers et al. 2013). Unless otherwise mentioned, all stimuli were presented at 65 dB (SPL), 32-bit depth and 400 kHz sample rate.

Electrophysiological recording. The electrodes were connected to the recording apparatus (Neuralynx digital Lynx) via a thin cable. The position of each tetrode was advanced by at least 40 μ m between sessions to avoid repeated recoding from the same units. Tetrode position was noted to ± 20 μ m precision. Electrophysiological data from 24 channels were filtered between 600 and 6,000 Hz (to obtain spike responses), digitized at 32 kHz, and stored for offline analysis. Single and multi-unit waveform clusters were isolated using commercial software (Plexon Spike Sorter) using previously described criteria (Carruthers et al. 2013).

Unit selection and firing-rate matching. To be included in analysis, a unit had to meet the following conditions: 1) its firing rate averaged at least 0.1 Hz firing rate during stimulus presentation, and 2) its spike count contained at least 0.78 bits/s of information about the vocalization identity during the presentation of at least one vocalization under one of the transformation conditions. We set this threshold to match the elbow in the histogram of the distribution of information rates for all recorded units that passed the firing rate threshold (see Fig. 5A, inset). We validated this threshold with visual inspection of vocalization response post-stimulus time histograms for units around the threshold. We estimated the information rate for each neuron by fitting a Poisson distribution to the distribution of spike counts evoked by each vocalization. We then computed the entropy of this set of 8 distributions, and subtracted from this value the prior entropy of 3 bits. Entropy was defined as $H(SR) = \sum_r p(r) H(SR = r) = - \sum_{r,s} p(r,s) \log_2 [p(s|r)]$. We defined $\pi_s(r) = \frac{\lambda_s^r}{r!} e^{-\lambda_s}$, the Poisson likelihood of detecting r spikes in response to stimulus s where λ_s is the mean number of spikes detected from a neuron in response to stimulus s . The entropy was computed as $H(SR) = - \frac{1}{N} \sum_{r,s} \Pi_s(r) \log_2 \left(\frac{\Pi_s(r)}{\sum_s \Pi_s(r)} \right)$. We performed

this computation separately for each transformation condition. In order to remove a potential source of bias due to different firing rate statistics in A1 and SRAF, we restricted all analyses to the subset of A1 units whose average firing rates most closely matched the selected SRAF units. We performed this restriction by recursively including the pair of units from the two areas with the most similar firing rates.

Response sparseness. To examine vocalization selectivity of recorded units, sparseness of vocalization was computed as:

$$\text{Sparseness} = 1 - \frac{\left(\sum_{i=1}^{i=n} FR_i/n \right)^2}{\sum_{i=1}^{i=n} FR_i^2/n}$$

where FR_i is the firing rate to vocalization i after the minimum firing rate in response to vocalizations was subtracted, and n is number of vocalizations included (which was 8). This value was computed separately for each recorded unit for each vocalization transformation, and then averaged over all transformations for recorded units from either A1 or SRAF.

Population response vector. The population response on each trial was represented as a vector, such that each element corresponded to responses of a unit to a particular presentation of a particular vocalization. Bin size for the spike count was selected by cross-validation (Hung et al. 2005; Rust and Dicarlo 2010); we tested classifiers using data binned at 50, 74, 100, and 150 ms. We found the highest performance in both A1 and SRAF when using a single bin 74 ms wide from vocalization onset, and we used this bin size for the remainder of the analyses. As each transformation of each vocalization was presented 100 times in each recording session, the analysis yielded $100 \times N$ matrix of responses for each of the 72 vocalization/transformation (8 vocalizations and 9 transformation conditions), where N was the number of units under analysis. The response of each unit was represented as an average of spike counts from 10 randomly selected trials. This pooling was performed after the segregation of vectors into training and validation data, such that the spike-counts used to produce the training data did not overlap with those used to produce the validation data.

Linear support vector machine (SVM) classifier. We used the support vector machine package *libsvm* (Chang and Lin 2011), as distributed by the *scikit-learn* project, version 0.15 (Pedregosa et al. 2011) to classify population response vectors. We used a linear kernel (resulting in decision boundaries defined by convex sets in the vector space of population spiking responses), and a soft-margin parameter of 1 (selected by cross-validation to maximize raw performance scores).

Classification procedure. For each classification task, a set of randomly selected N units (unless otherwise noted, we used $N = 60$) was used to construct the population response vector as described above, dividing the data into training and validation sets. For each vocalization, 80 vectors were used to train and 20 to validate per-transformation and within-transformation classification (see *Across-transformation performance* below). In order to divide the data evenly among the nine transformations, 81 vectors were used to train and 18 to validate in all-transformation classification. We used the vectors in the training dataset to fit a classifier, and then tested the ability of the resulting classifier to determine which of the vocalizations evoked each of the vectors in the validation dataset.

Bootstrapping. The entire classification procedure was repeated 1000 times for each task, each time on a different randomly selected

Fig. 1. Spectrograms of vocalizations and transformations used as acoustic stimuli in the experiments. A: the eight different original vocalizations selected from recordings, after de-noising. B: one original vocalization (center), as well as the 8 different transformations of that vocalization presented in the experiment. From top left to bottom right: T+: temporally stretched by factor of 1.25; CF+: center frequency shifted up to 7.9 kHz; T-: temporally compressed by factor of 0.75; FM-: frequency modulation scaled by a factor of 0.75; Original: denoised original vocalization; FM+: frequency modulation scaled by a factor of 1.25; T-/FM-: temporally compressed and frequency modulation scaled by a factor of 0.75; CF-: center frequency shifted down by 7.9 kHz; T+/FM+: temporally stretched and frequency modulation scaled by a factor of 1.25.

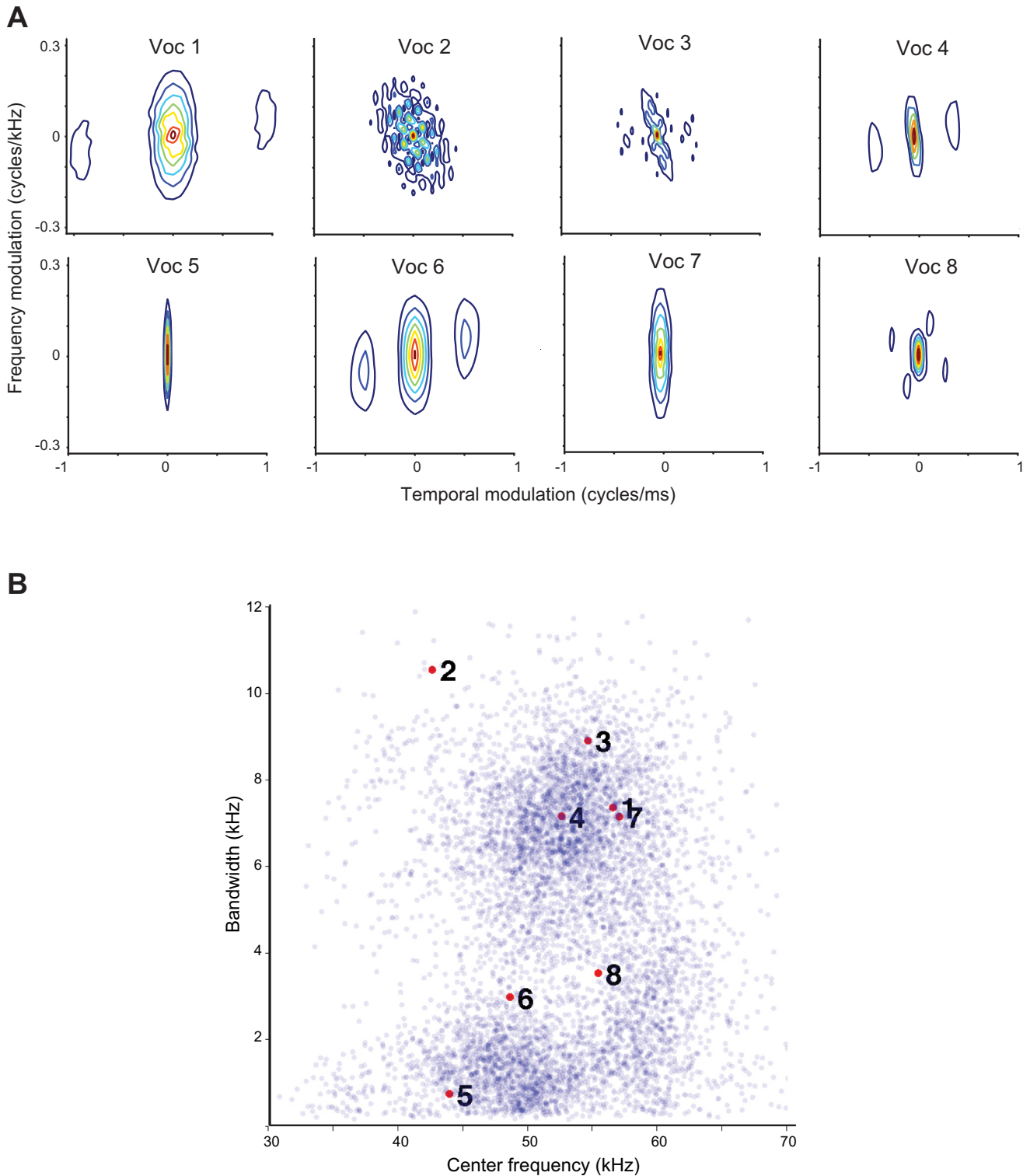


Fig. 2. Statistical characterization of vocalizations. *A*: spectrotemporal modulation spectrum for the 8 vocalizations. *B*: distribution of center frequency and bandwidth for all recorded vocalizations. Eight vocalizations used in the study are indicated by red dots with corresponding numbers.

population of units, and each time using a different randomly selected set of trials for validation.

Mode of classification. Classification was performed in one of two modes: In the pairwise mode, we trained a separate binary classifier for each possible pair of vocalizations, and classified which of the two vocalizations evoked each vector. In one-vs.-all

mode, we trained an 8-way classifier on responses to all vocalizations at once, and classified which of the eight vocalizations was most likely to evoke each response vector (Chang and Lin 2011; Pedregosa et al. 2011). This was implemented by computing all pairwise classifications followed by a voting procedure. We recorded the results of each classification, and computed the perfor-

mance of the classifier as the fraction of response vectors that it classified correctly. As there were 8 vocalizations, performance was compared to the chance value of 0.125 in one-vs.-all mode and to 0.5 in pairwise mode.

Across-transformation performance. We trained and tested classifiers on vectors drawn from a subset of different transformation conditions. We chose the subset of transformations in two different ways: When testing per-transformation performance, we trained and tested on vectors drawn from presentations of one transformation and from the original vocalizations. When testing all-transformation performance, we trained and tested on vectors drawn from all 9 transformation conditions.

Within-transformation performance. For each subset of transformations on which we tested across-transformation performance, we also trained and tested classifiers on responses under each individual transformation condition. We refer to performance of these classifiers, averaged over the transformation conditions, as the within-transformation performance.

Generalization penalty. In order to evaluate how tolerant neural codes are to stimulus transformation, we compared the performance on generalization tasks with the performance on the corresponding within-transformation tasks. We defined the generalization penalty as the difference between the within- and across-transformation performance.

RESULTS

In order to measure how invariant neural population responses to vocalizations are to their acoustic transformations, we selected USV exemplars and constructed their transformations along basic acoustic dimensions. Rat USVs consist of frequency modulated pure tones with little or no harmonic structure. The simple structure of these vocalizations makes it possible to extract the vocalization itself from background noise with high fidelity. Their simplicity also allows us to parameterize the vocalizations; they are characterized by the dominant frequency, and the amplitude at that frequency, as these quantities vary with time. In turn, this simple parameterization allows us to easily and efficiently transform aspects of the vocalizations. The details of this parameterization and transformation process are reported in depth in our previously published work (Carruthers et al. 2013).

We selected 8 distinct vocalizations from recordings of social interactions between male adolescent rats (Carruthers et al. 2013; Sirotin et al. 2014). We chose these vocalizations to include a variety of temporal and frequency modulation spectra (Fig. 2A) and to cover the center frequency and frequency bandwidth distribution of the full set of recorded vocalizations (Fig. 2B). We previously demonstrated that the responses of neurons to vocalizations were dominated by modulation in frequency and amplitude (Carruthers et al. 2013). Therefore, we used frequency, frequency modulation, and amplitude modulation time course as the relevant acoustic dimensions to generate transformed vocalizations. We constructed 8 different transformed versions of these vocalizations by adjusting the center frequency, duration, and/or spectral bandwidth of these vocalizations (see METHODS), for a total of 9 versions of each vocalization. The 8 original vocalizations we selected can be seen in Fig. 1A, and Fig. 1B shows the different transformed versions of *vocalization 3*. We recorded neural responses in A1 and SRAF in rats as they passively listened to these original and transformed vocalizations. As in our previous study (Carruthers et al. 2013), we found that A1 units respond selectively and with high temporal precision to USVs (Fig. 3). SRAF units

exhibited similar patterns of responses (Fig. 4). For instance, the representative A1 unit shown in Fig. 3 responded significantly to all of the original vocalizations except *vocalizations 4, 6, and 7 (row 1)*. Meanwhile, the representative SRAF unit in Fig. 4 responded significantly to all of the original vocalizations except *vocalization 6 (row 1)*. Note that the A1 unit's response to *vocalization 5* varies significantly in both size and temporal structure when the vocalization is transformed. Meanwhile, the SRAF unit's response to the same vocalization is consistent regardless of which transformation of the vocalization is played. In this instance, the selected SRAF unit exhibits greater invariance to transformations of *vocalization 5* than the selected A1 unit.

To compare the responses of populations of units in A1 and SRAF and to ensure that the effects that we observe are not due simply to increased information capacity of neurons that fire at higher firing rates, we selected subpopulations of units that were matched for firing rate distribution (Rust and Dicarlo 2010; Ulanovsky et al. 2004) (Fig. 5A). We then compared the tuning properties of units from the two brain areas, as measured by the pure-tone frequency that evoked the highest firing rate from the units. We found no difference in the distribution of best frequencies between the two populations (Kolmogorov-Smirnov test, $P = 0.66$) (Fig. 5B). We compared the amount of information transmitted about a vocalization's identity by the spike counts of units in each brain area, and again found no significant difference (Fig. 5C, Kolmogorov-Smirnov test, $P = 0.42$). Furthermore, we computed sparseness of responses of A1 and SRAF units to vocalizations, which is a measure of neuronal selectivity to vocalizations. A sparseness value of 1 indicates that the unit responds differently to a single vocalization than to all others, whereas a sparseness value of 0 indicates that the unit responds equally to all vocalizations. The mean sparseness values for responses were 0.354 for A1, and 0.376 for SRAF (Fig. 5D), but this difference was not significant (Kolmogorov-Smirnov test, $P = 0.084$). These analyses demonstrate that the selected neuronal populations in A1 and SRAF were similarly selective to vocalizations.

Neuronal populations in A1 and SRAF exhibited similar performance in their ability to classify responses to different vocalizations. We trained classifiers to distinguish between original vocalizations on the basis of neuronal responses, and we measured the resulting performances. To ensure that the results were not skewed by a particular vocalization, we computed the classification either for responses to each pair of vocalizations (pairwise performance), or for responses to all 8 vocalizations simultaneously (8-way performance). We found a small but significant difference between the average performance of those classifiers trained and tested on A1 responses and those trained and tested on the SRAF responses (Fig. 5, E and F), but the results were mixed. Pairwise classifications performed on populations of A1 units were 88.0% correct, and on populations of SRAF units, 88.5% correct (Kolmogorov-Smirnov test, $P = 0.0013$). On the other hand, 8-way classifications performed on populations of 60 A1 units were 61% correct, and on SRAF units were 59% correct (Kolmogorov-Smirnov test, $P = 7.7e-11$). Figure 5, G and H, shows the classification performance broken down by vocalization for pairwise classification for A1 (Fig. 5G) and SRAF (Fig. 5H). There is high variability in performance between vocalization pairs for either brain area. However, the performance levels are



Fig. 3. Peristimulus-time raster plots (above) and histograms (below) of an exemplar A1 unit showing selective responses to vocalization stimuli. Each column corresponds to one original vocalization, and every two rows to one transformation of that vocalization. Histograms were first computed for 1-ms time bins, and then smoothed with 11-ms Hanning window.

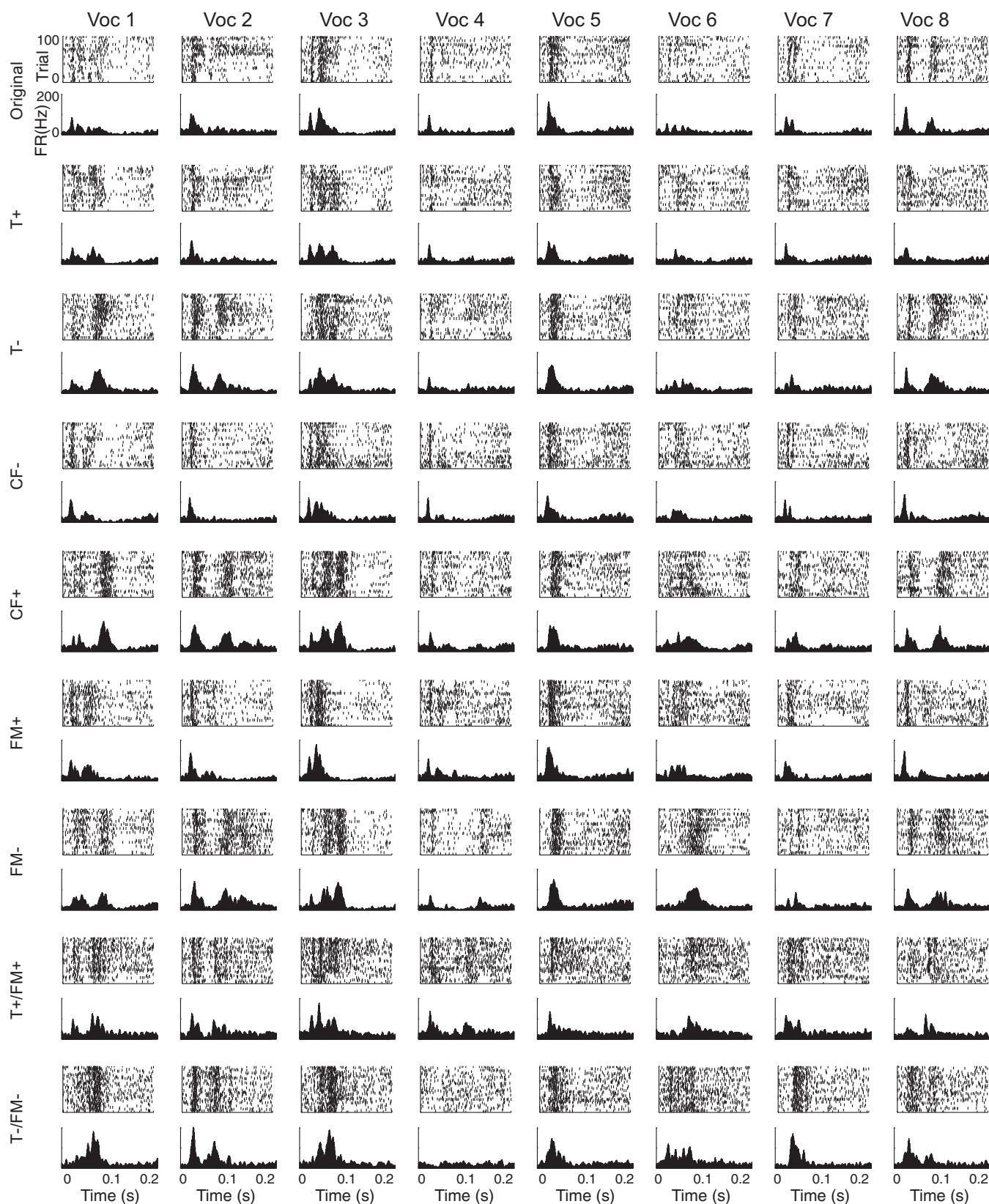
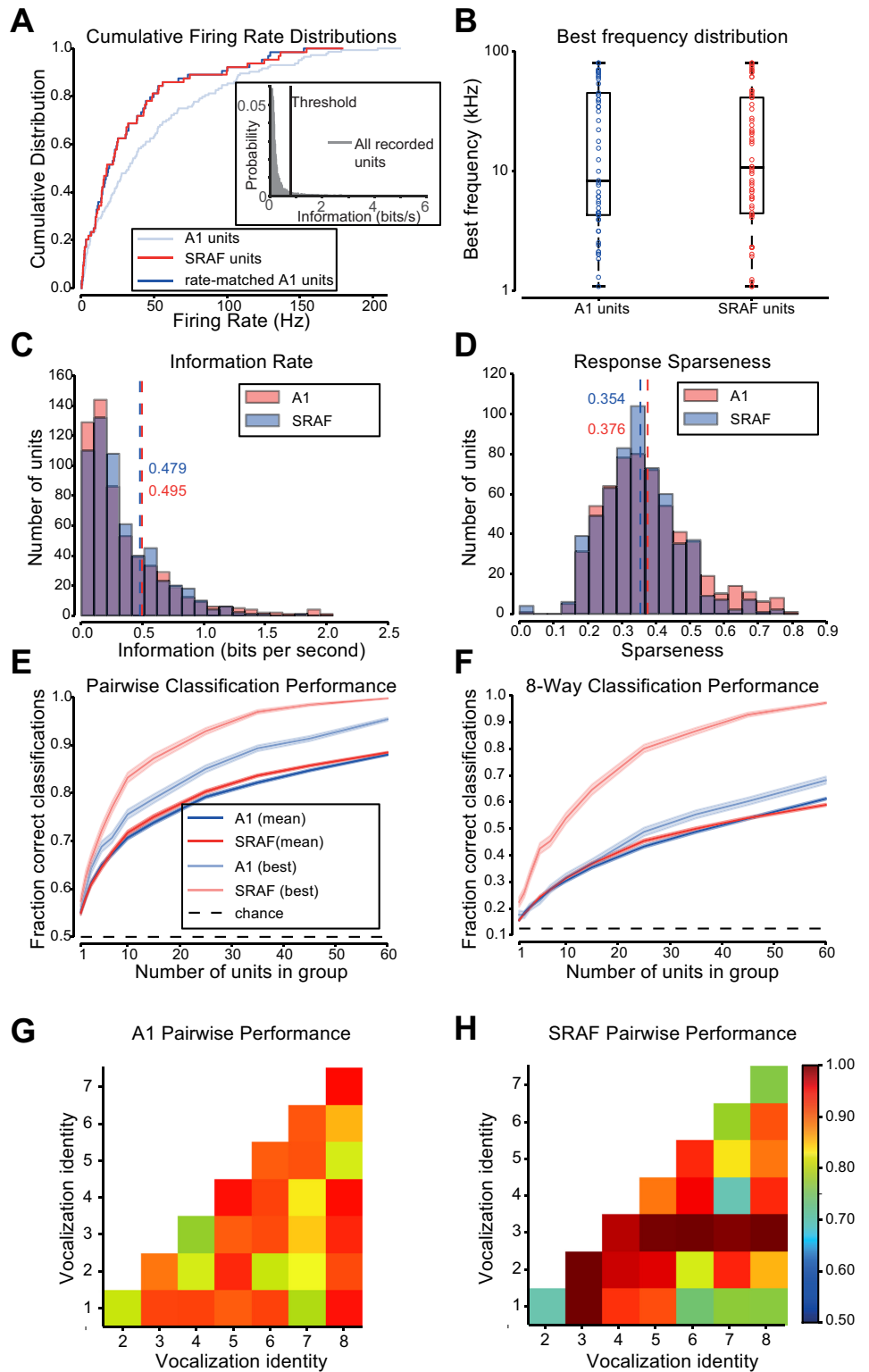


Fig. 4. Peristimulus-time raster plots (above) and histograms (below) of an exemplar SRAF unit showing selective responses to vocalization stimuli. Each column corresponds to one original vocalization, and every two rows to one transformation of that vocalization. Histograms were first computed for 1 ms time-bins, and then smoothed with 11-ms Hanning window.

Fig. 5. Ensembles of A1 and suprarhinal auditory field (SRAF) units under study are similar in responses and overall classification performance. **A:** cumulative distributions for average firing rate of units during stimulus presentation. Distribution of SRAF units shown in red, A1 units shown in faint blue, and the subset of A1 units matched to the SRAF units shown in blue. *Inset:* distribution of information rates for all recorded units that passed the minimum firing rate criterion. The threshold for information rate (0.78 bits/s) in response to vocalizations under at least one transformation is marked by a vertical black line. **B:** box-plot showing the distribution of frequency tunings of the units selected from A1 and from SRAF. The boxes show the extent of the central 50% of the data, with the horizontal bar showing the median frequency. **C:** histogram of the information contained in the spike counts of units from A1 and SRAF about each vocalization. Dashed lines mark the mean values. **D:** histogram of sparseness (with respect to vocalization identity) of responses of units from A1 and SRAF. Dashed lines mark the mean values. **E:** classification accuracy of support vector machine (SVM) classifier distinguishing between two vocalizations (pairwise mode). Faded colors show performance for the pair of vocalizations with the highest performance for each brain area, and saturated colors show average performance across pairs. **F:** classification accuracy of SVM classifier distinguishing between all vocalizations (8-way mode). Faded colors show performance for the vocalization with the highest performance for each brain area, and saturated colors show average performance across all vocalizations. **G:** average performance of pairwise classification for each vocalization for neuronal populations in A1. **H:** average performance of pairwise classification for each vocalization for neuronal populations in SRAF.



similar. Together, these results indicate that neuronal populations in A1 and SRAF are similar in their ability to classify vocalizations.

To test whether neuronal populations exhibited invariance to transformations in classifying vocalizations, we measured whether the ability of neuronal populations to classify vocalizations was reduced when vocalizations were distorted acous-

tically. Therefore, we trained and tested classifiers for vocalizations based on population neuronal responses and compared their performance under within-transformation and across-transformation conditions (Fig. 6A). In within-transformation condition, the classifiers were trained and tested to discriminate responses to vocalizations under a single transformation. In across-transformation condition, the classifier was trained and

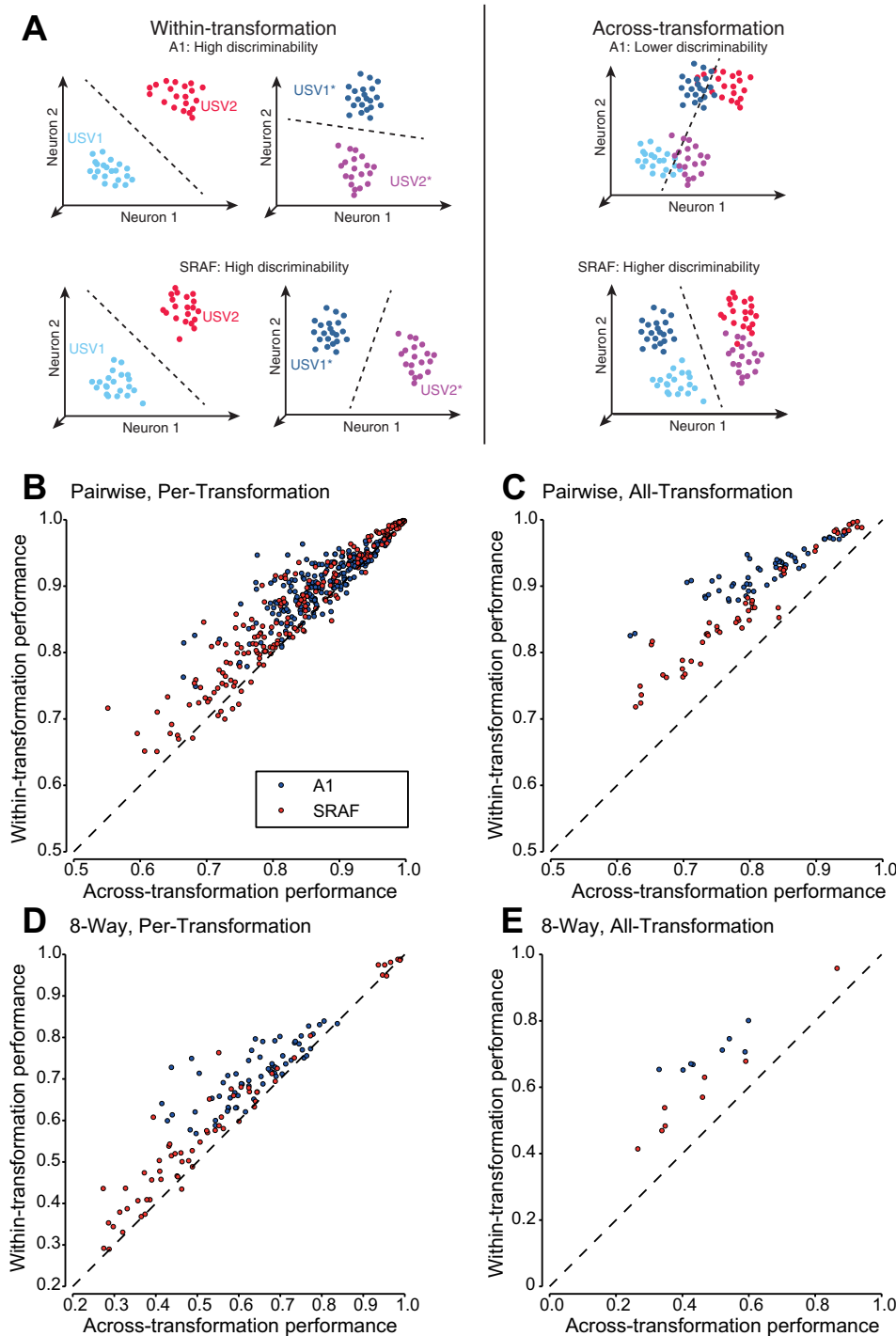


Fig. 6. Classifier performance on within-transformation and across-transformation conditions. *A*: schematic diagram of neuronal responses to 2 original (USV1, USV2) and transformed (USV1*, USV2*) vocalizations. Each dot denotes a population response vector projected in a low-dimensional subspace. *Left*: within-transformation classification: classifier is trained and tested to classify responses to vocalizations for a single transformation. Within-transformation discriminability is high for both original and transformed vocalizations by populations of neurons in either A1 (top) or SRAF (bottom). *Right*: generalization classification: Classifier is trained and tested to classify responses to original and transformed vocalizations simultaneously. Predictions of the hierarchical coding model: Across-transformation classification performance is low for A1 and high for SRAF, reflecting an increase in invariance from A1 to SRAF. *B* and *C*: performance when discriminating each vocalization from one other vocalization (pairwise classification). *D* and *E*: performance when discriminating each vocalization from all others (8-way classification). *B* and *D*: performance when generalization is performed across the original vocalizations and one transformation at a time (per-transformation). *C* and *E*: performance when generalization is performed across all eight transformations and the originals at once (all-transformation).

tested in discriminating responses to vocalizations in original form and one or all transformations. The difference between within-transformation and the across-transformation classifier performance was termed the generalization penalty. If the neuronal population exhibited low invariance, we expected the across-transformation performance to be lower than within-transformation performance and the generalization penalty to be high (Fig. 6*A*, top). If neuronal population exhibited high invariance, we expected the across-transformation performance to be equal to within-transformation performance and the generalization penalty to be low (Fig. 6*A*, bottom).

To ensure that responses to a select transformation were not skewing the results, we computed across-transformation performance both for each of the transformations and for all transformations. In per-transformation condition, the classifier was trained and tested in discriminating responses to vocalizations in original form and under one other transformation. In all-transformation condition, the classifier was trained and tested in discrimination of responses to vocalizations in original form and under all 8 transformations simultaneously.

Neuronal populations in A1 exhibited greater reduction in performance on across-transformation condition compared to

within-transformation condition than neuronal population in SRAF. Figures 6 and 7 present the comparison between across-transformation performance and within-transformation performance for each of the different conditions. Note that the different conditions result in very different numbers of data points: the per-transformation conditions have 8 times as many data points as the all-transformation conditions, as the former yields a separate data point for each transformation. Similarly, the pairwise conditions yield 28 times as many data points as the 8-way conditions (one for each unique pair drawn from the

8 vocalizations). As expected, for both A1 and SRAF, the classification performance was higher for within-transformation than across-transformation condition (Fig. 6, *B–E*). However, the difference in performance between within-transformation and across-transformation conditions was higher in A1 than in SRAF: SRAF populations suffered a smaller generalization penalty under all conditions tested (Fig. 7), indicating that neuronal ensembles in SRAF exhibited greater generalization than in A1. This effect was present under both pairwise (Fig. 6, *B* and *C*, and Fig. 7, *A* and *B*) and 8-way classification

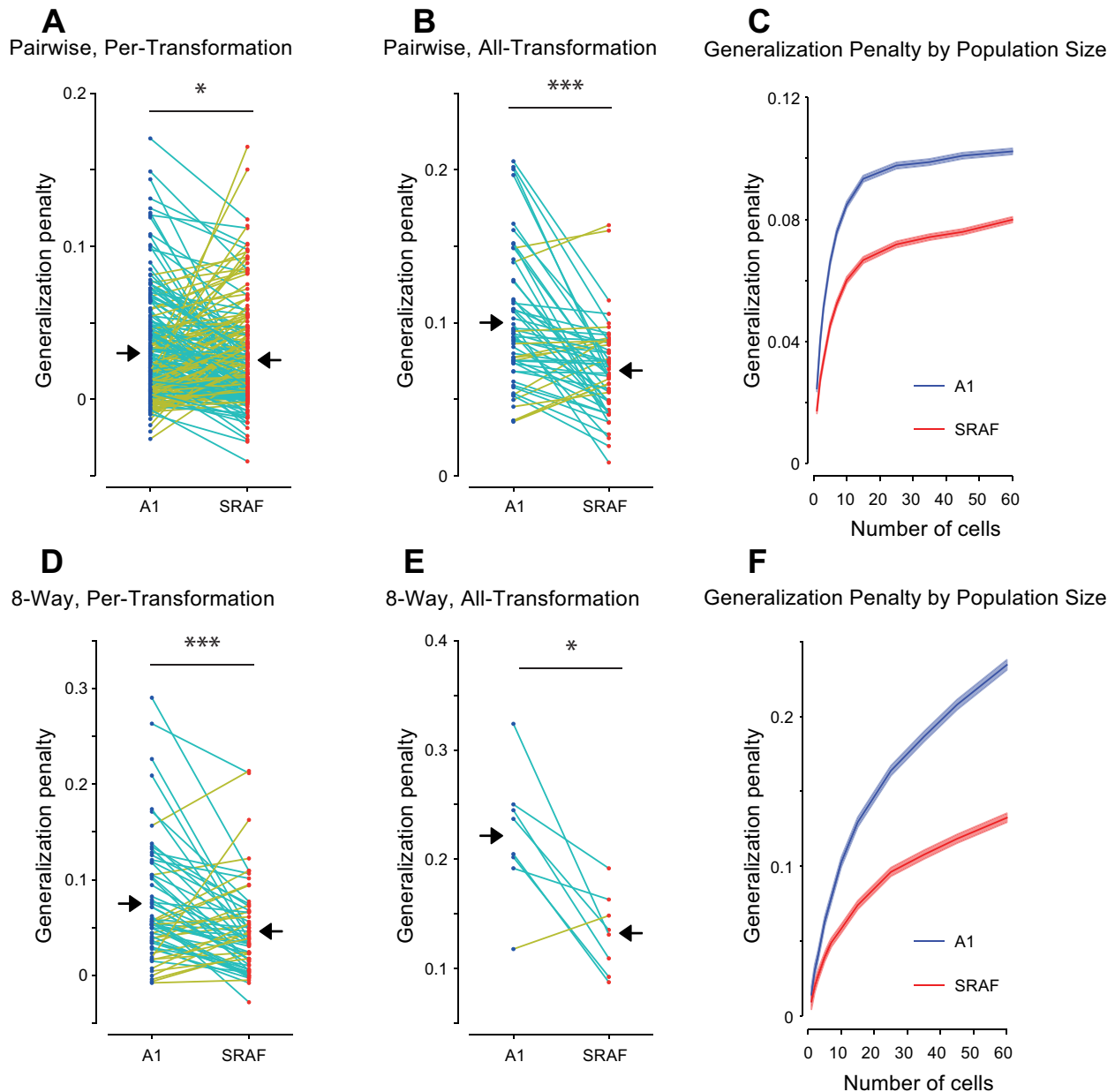


Fig. 7. Generalization penalty (difference between within-transformation performance and across-transformation performance) is higher for A1 ensembles than for SRAF ensembles. Each dot corresponds to average classifier performance for a specific vocalization/transformation combination. Conditions in which SRAF units show smaller penalty than A1 units are connected with cyan lines, conditions, in which SRAF units show more penalty are connected by yellow lines. Mean penalty values for each brain area are marked with black arrows. *A* and *B*: generalization penalty when discriminating each vocalization from one other vocalization (pairwise classification). *D* and *E*: generalization penalty when discriminating each vocalization from all others (8-way classification). *A* and *D*: generalization penalty when generalization is performed only across the original vocalizations and one vocalization at a time (per-transformation generalization). *B* and *E*: generalization penalty when generalization is performed across all eight transformations and the originals at once (all-transformation generalization). *C* and *F*: generalization penalty as function of the number of cells in ensemble. *C*: pairwise classification across all eight transformations, as in *B*. *E*: 8-way classification across eight transformations, as in *D*. * $P < 0.05$; *** $P < 0.001$.

(Fig. 6, *D* and *E*, and Fig. 7, *C* and *D*), and for generalization in per-transformation (Fig. 6, *B* and *D*, and Fig. 7, *A* and *D*, pairwise classification, $P = 0.028$; 8-way classification, $P = 1.9e-4$; Wilcoxon paired sign-rank test; 60 units in each ensemble tested) and all-transformation mode (Fig. 6, *C* and *E*, and Fig. 7, *B* and *E*; pairwise classification, $P = 1.4e-5$; 8-way classification, $P = 0.025$; Wilcoxon paired sign-rank test; 60 units in each ensemble tested). The greater generalization penalty for A1 as compared to SRAF was preserved for increasing number of neurons in the ensemble, as the discrimination performance improved and the relative difference between across- and within- performance increased (Fig. 7, *C* and *F*). Taken together, we find that populations of SRAF units are better able to generalize across acoustic transformations of stimuli than populations of A1 units, as characterized by linear encoding of stimulus identity. These results suggest that populations of SRAF neurons are more invariant to transformations of auditory objects than populations of A1 neurons.

DISCUSSION

Our goal was to test whether and how populations of neurons in the auditory cortex represented vocalizations in an invariant fashion. We tested whether neurons in the nonprimary area SRAF exhibit greater invariance to simple acoustic transformations than do neurons in A1. To estimate invariance in neuronal encoding of vocalizations, we computed the difference in the ability of neuronal population codes to classify vocalizations between different types following acoustic distortions of vocalizations (Fig. 1). We found that, while neuronal populations in A1 and SRAF exhibited similar selectivity to vocalizations (Figs. 3–5), neuronal populations in SRAF exhibited higher invariance to acoustic transformations of vocalizations than in A1, as measured by lower generalization penalty (Figs. 6 and 7). These results are consistent with the hypothesis that invariance arises gradually within the auditory pathway, with higher auditory areas exhibiting progressively higher invariances toward basic transformations of acoustic signals. An invariant representation at the level of population neuronal ensemble activity supports the ability to discriminate between behaviorally important sounds (such as vocalizations and speech) despite speaker variability and environmental changes.

We recently found that rat ultrasonic vocalizations can be parameterized as amplitude- and frequency-modulated tones, similar to whistles (Carruthers et al. 2013). Units in the auditory cortex exhibited selective responses to subsets of the vocalizations, and a model that relies on the amplitude- and frequency-modulation time course of the vocalizations could predict the responses to novel vocalizations. These results point to amplitude- and frequency modulations as essential acoustic dimensions for encoding of ultrasonic vocalizations. Therefore, in this study, we tested four types of acoustic distortions based on basic transformations of these dimensions: temporal dilation, frequency shift, frequency modulation scaling and combined temporal dilation and frequency modulation scaling. These transformations likely carry behavioral significance and might be encountered when a speaker's voice is temporally dilated, or be characteristic of different speakers (Fitch et al. 1997). While there is limited evidence that such transformations are typical in vocalizations emitted by rats, preliminary analysis of rat vocalizations revealed a large range of variability in these parameters across vocalizations.

Neurons throughout the auditory pathway have been shown to exhibit selective responses to vocalizations. In response to ultrasonic vocalizations, neurons in the auditory midbrain exhibit a mix of selective and nonselective responses in rodents (Holmstrom et al. 2010; Pincherli Castellanos et al. 2007). At the level of A1, neurons across species respond strongly to conspecific vocalizations (Gehr et al. 2000; Glass and Wollberg 1983; Huetz et al. 2009; Medvedev and Kanwal 2004; Pelleg-Toiba and Wollberg 1991; Wallace et al. 2005; Wang et al. 1995). The specialization of neuronal responses for the natural statistics of vocalization has been under debate (Huetz et al. 2009; Wang et al. 1995). The avian auditory system exhibits strong specialization for natural sounds and conspecific vocalizations (Schneider and Woolley 2010; Woolley et al. 2005), and a similar hierarchical transformation has been observed between primary and secondary cortical analogs (Elie and Theunissen 2015). In rodents, specialized responses to USVs in A1 are likely context-dependent (Galindo-Leon et al. 2009; Liu et al. 2006; Liu and Schreiner 2007; Marlin et al. 2015). Therefore, extending our study to be able to manipulate the behavioral “meaning” of the vocalizations through training will greatly enrich our understanding of how the transformation that we observe contributes to auditory behavioral performance.

A1 neurons adapt to the statistical structure of the acoustic stimulus (Asari and Zador 2009; Blake and Merzenich 2002; Kvale and Schreiner 2004; Natan et al. 2015; Rabinowitz et al. 2013; Rabinowitz et al. 2011). The amplitude of frequency shift and frequency modulation scaling coefficient were chosen on the basis of the range of the statistics of ultrasonic vocalizations that we recorded (Carruthers et al. 2013). These manipulations were designed to keep the statistics of the acoustic stimulus within the range of original vocalizations, in order to best drive responses in A1. Psychophysical studies in humans found that speech comprehension is preserved over temporal dilations up to a factor of 2 (Beasley et al. 1980; Dupoux and Green 1997; Foulke and Sticht 1969). Here, we used a scaling factor of 1.25 or 0.75, similar to previous electrophysiological studies (Gehr et al. 2000; Wang et al. 1995), and also falling within the statistical range of the recorded vocalizations. Furthermore, we included a stimulus in which frequency modulation scaling was combined with temporal dilation. This transformation was designed in order to preserve the velocity of frequency modulation from the original stimulus. The observed results exhibit robustness to the type of transformation that was applied to the stimulus, and are therefore likely generalizable to transformations of other acoustic features.

In order to quantify the invariance of population neuronal codes, we used the performance of automated classifiers as a lower bound for the information available in the population responses to original and transformed vocalizations. To assay generalization performance, we computed the difference between classifier performance on within- and across-transformation conditions. We expected this difference to be small for populations of neurons that generalized, and large for populations of neurons that did not exhibit generalization (Fig. 6A). Computing this measure was particularly important, as populations of A1 and SRAF neurons exhibited a great degree of variability in classification performance for both within- and across-transformation classification (Fig. 6, *B–E*). This variability is consistent

with the known details about heterogeneity in neuronal cell types and connectivity in the mammalian cortex (Kanold et al. 2014). Therefore, measuring the relative improvement in classification performance using the generalization penalty overcomes the limits of heterogeneity in performance.

In order to probe the transformation of representations from one brain area to the next, we decided to limit the classifiers to information that could be linearly decoded from population responses. For this reason, we chose to use linear support vector machines (SVMs, see METHODS) for classifiers. SVMs are designed to find robust linear boundaries between classes of vectors in a high-dimensional space. When trained on two sets of vectors, an SVM finds a hyperplane (a flat, infinite boundary) that provides the best separation between the two sets: a hyperplane that divides the space in two, assigning every vector on one side to the first set, and everything on the other side to the second. In this case finding the “best separation” means a trade-off between having as many of the training vectors as possible be on the correct side, and giving the separating hyperplane as large of a margin (the distance between the hyperplane and the closest correctly classified vectors) as possible (Dayan and Abbott 2005; Vapnik 2000). The result is generally a robust, accurate decision boundary that can be used to classify a vector into one of the two sets. A linear classification can be viewed as a weighted summation of inputs, followed by a thresholding operation, a combination of actions that is understood to be one of the most fundamental computations performed by neurons in the brain (Abbott 1994; deCharms and Zador 2000). Therefore, examination of information via linear classifiers places a lower bound on the level of classification that could be accomplished during the next stage of neural processing.

Several mechanisms could potentially explain the increase in invariance we observe between A1 and SRAF. As previously suggested, cortical microcircuits in A1 can transform incoming responses into a more feature-invariant form (Atencio et al. 2009). By integrating over neurons with different tuning properties, higher level neurons can develop tuning to more specific conjunction of features (becoming more selective), while exhibiting invariance to basic transformations. Alternatively, higher auditory brain areas may be better able to adapt to the basic statistical features of auditory stimuli, such that the neuronal responses would be sensitive to patterns of spectrotemporal modulation regardless of basic acoustic transformations. At the level of the midbrain, adaptation to the stimulus variance allows for invariant encoding of stimulus amplitude fluctuations (Rabinowitz et al. 2013). In the mouse inferior colliculus, neurons exhibit heterogeneous response to ultrasonic vocalizations and their acoustically distorted versions (Holmstrom et al. 2010). At higher processing stages, as auditory processing becomes progressively multidimensional (Sharpee et al. 2011), adaptation could produce a neural code that could be more robustly decoded across stimulus transformations. More complex population codes may provide a greater amount of information in the brain (Averbeck et al. 2006; Averbeck and Lee 2004; Cohen and Kohn 2011). Extensions to the present study could be used to distinguish between invariance due to statistical adaptation, and invariance due to feature independence in neural responses.

While our results support a hierarchical coding model for the representation of vocalizations across different stages of the au-

ditory cortex, the observed changes may originate at the subcortical level, e.g., inferior colliculus (Holmstrom et al. 2010) or differential thalamocortical inputs (Covic and Sherman 2011), and already should be encoded within specific groups of neurons or within different cortical layers within the primary auditory cortex. Further investigation including more selective recording and targeting of specific cell types is required to pinpoint whether the transformation occurs throughout the pathway or within the canonical cortical circuit.

ACKNOWLEDGMENTS

We thank Drs. Y. Cohen, S. Eliades, M. Pagan, and N. Rust and members of the Geffen laboratory for helpful discussions on analysis, and L. Liu, L. Cheung, A. Davis, A. Nguyen, A. Chen, and D. Mohabir for technical assistance with experiments.

GRANTS

This work was supported by NIDCD Grants R03-DC-013660 and R01-DC-014479, Klingenstein Foundation Award in Neurosciences, Burroughs Wellcome Fund Career Award at Scientific Interface, Human Frontiers in Science Foundation Young Investigator Award and Pennsylvania Lions Club Hearing Research Fellowship to M. N. Geffen. I. M. Carruthers and A. Jaegle were supported by the Cognition and Perception IGERT training grant. A. Jaegle was also partially supported by the Hearst Foundation Fellowship. J. J. Briguglio was partially supported by NSF PHY1058202 and US-Israel BSF 2011058.

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the author(s).

AUTHOR CONTRIBUTIONS

Author contributions: I.M.C., D.A.L., A.J., and M.N.G. conception and design of research; I.M.C., D.A.L., A.J., L.M.-T., R.G.N., and M.N.G. performed experiments; I.M.C., D.A.L., A.J., J.J.B., L.M.-T., R.G.N., and M.N.G. analyzed data; I.M.C., D.A.L., and M.N.G. interpreted results of experiments; I.M.C., D.A.L., J.J.B., and M.N.G. prepared figures; I.M.C. and M.N.G. drafted manuscript; I.M.C., D.A.L., A.J., J.J.B., L.M.-T., R.G.N., and M.N.G. approved final version of manuscript; D.A.L., A.J., J.J.B., R.G.N., and M.N.G. edited and revised manuscript.

REFERENCES

- Abbott LF.** Decoding neuronal firing and modelling neural networks. *Q Rev Biophys* 27: 291–331, 1994.
- Aizenberg M, Geffen MN.** Bidirectional effects of auditory aversive learning on sensory acuity are mediated by the auditory cortex. *Nat Neurosci* 16: 994–996, 2013.
- Arnault P, Roger M.** Ventral temporal cortex in the rat: connections of secondary auditory areas Te2 and Te3. *J Comp Neurol* 302: 110–123, 1990.
- Asari H, Zador A.** Long-lasting context dependence constrains neural encoding models in rodent auditory cortex. *J Neurophysiol* 102: 2638–2656, 2009.
- Atencio C, Sharpee T, Schreiner C.** Hierarchical computation in the canonical auditory cortical circuit. *Proc Natl Acad Sci USA* 106: 21894–21899, 2009.
- Averbeck BB, Latham PE, Pouget A.** Neural correlations, population coding and computation. *Nat Rev Neurosci* 7: 358–366, 2006.
- Averbeck BB, Lee D.** Coding and transmission of information by neural ensembles. *Trends Neurosci* 27: 225–230, 2004.
- Barfield RJ, Auerbach P, Geyer LA, McIntosh TK.** Ultrasonic vocalizations in rat sexual-behavior. *Am Zool* 19: 469–480, 1979.
- Beasley DS, Bratt GW, Rintelmann WF.** Intelligibility of time-compressed sentential stimuli. *J Speech Hear Res* 23: 722–731, 1980.
- Bialy M, Rydz M, Kaczmarek L.** Precontact 50-kHz vocalizations in male rats during acquisition of sexual experience. *Behav Neurosci* 114: 983–990, 2000.
- Bizley JK, Walker KM, Silverman BW, King AJ, Schnupp JW.** Interdependent encoding of pitch, timbre, and spatial location in auditory cortex. *J Neurosci* 29: 2064–2075, 2009.

- Blake DT, Merzenich MM.** Changes of AI receptive fields with sound density. *J Neurophysiol* 88: 3409–3420, 2002.
- Brudzynski SM.** Communication of adult rats by ultrasonic vocalization: biological, sociobiological, and neuroscience approaches. *ILAR J* 50: 43–50, 2009.
- Brudzynski SM, Pniak A.** Social contacts and production of 50-kHz short ultrasonic calls in adult rats. *J Comp Psychol* 116: 73–82, 2002.
- Burgdorf J, Knutson B, Panksepp J.** Anticipation of rewarding electrical brain stimulation evokes ultrasonic vocalization in rats. *Behavioral Neurosci* 114: 320–327, 2000.
- Burgdorf J, Kroes RA, Moskal JR, Pfaus JG, Brudzynski SM, Panksepp J.** Ultrasonic vocalizations of rats (*Rattus norvegicus*) during mating, play, and aggression: Behavioral concomitants, relationship to reward, and self-administration of playback. *J Comp Psychol* 122: 357–367, 2008.
- Carrasco A, Lomber SG.** Neuronal activation times to simple, complex, and natural sounds in cat primary and nonprimary auditory cortex. *J Neurophysiol* 106: 1166–1178, 2011.
- Carruthers IM, Natan RG, Geffen MN.** Encoding of ultrasonic vocalizations in the auditory cortex. *J Neurophysiol* 109: 1912–1927, 2013.
- Chang CC, Lin CJ.** LIBSVM: A Library for Support Vector Machines. *Acm T Intel Syst Tec* 2: 2011.
- Cohen MR, Kohn A.** Measuring and interpreting neuronal correlations. *Nat Neurosci* 14: 811–819, 2011.
- Covic EN, Sherman SM.** Synaptic properties of connections between the primary and secondary auditory cortices in mice. *Cereb Cortex* 21: 2425–2441, 2011.
- Dayan P, Abbott LF.** *Theoretical Neuroscience Computational and Mathematical Modeling of Neural Systems.* Cambridge, MA: MIT Press, 2005.
- deCharms RC, Zador A.** Neural representation and the cortical code. *Annu Rev Neurosci* 23: 613–647, 2000.
- DiCarlo JJ, Cox DD.** Untangling invariant object recognition. *Trends Cogn Sci* 11: 333–341, 2007.
- Dupoux E, Green K.** Perceptual adjustment to highly compressed speech: effects of talker and rate changes. *J Exp Psychol Hum Percept Perform* 23: 914–927, 1997.
- Elie JE, Theunissen FE.** Meaning in the avian auditory cortex: neural representation of communication calls. *Eur J Neurosci* 41: 546–567, 2015.
- Engineer CT, Perez CA, Chen YH, Carraway RS, Reed AC, Shetake JA, Jakkamsetti V, Chang KQ, Kilgard MP.** Cortical activity patterns predict speech discrimination ability. *Nat Neurosci* 11: 603–608, 2008.
- Fitch RH, Miller S, Tallal P.** Neurobiology of speech perception. *Annu Rev Neurosci* 20: 331–353, 1997.
- Foulke E, Sticht TG.** Review of research on the intelligibility and comprehension of accelerated speech. *Psychol Bull* 72: 50–62, 1969.
- Fritz JB, David SV, Radtke-Schuller S, Yin P, Shamma SA.** Adaptive, behaviorally gated, persistent encoding of task-relevant auditory information in ferret frontal cortex. *Nat Neurosci* 13: 1011–1019, 2010.
- Gai Y, Carney LH.** Statistical analyses of temporal information in auditory brainstem responses to tones in noise: correlation index and spike-distance metric. *J Assoc Res Otolaryngol* 9: 373–387, 2008.
- Galindo-Leon EE, Lin FG, Liu RC.** Inhibitory plasticity in a lateral band improves cortical detection of natural vocalizations. *Neuron* 62: 705–716, 2009.
- Gehr DD, Komiya H, Eggermont JJ.** Neuronal responses in cat primary auditory cortex to natural and altered species-specific calls. *Hear Res* 150: 27–42, 2000.
- Glass I, Wollberg Z.** Responses of cells in the auditory cortex of awake squirrel monkeys to normal and reversed species-specific vocalizations. *Hear Res* 9: 27–33, 1983.
- Holmstrom LA, Eeuwes LB, Roberts PD, Portfors CV.** Efficient encoding of vocalizations in the auditory midbrain. *J Neurosci* 30: 802–819, 2010.
- Huetz C, Philibert B, Edeline JM.** A spike-timing code for discriminating conspecific vocalizations in the thalamocortical system of anesthetized and awake guinea pigs. *J Neurosci* 29: 334–350, 2009.
- Hung CP, Kreiman G, Poggio T, DiCarlo JJ.** Fast readout of object identity from macaque inferior temporal cortex. *Science* 310: 863–866, 2005.
- Kaas JH, Hackett TA.** Subdivisions of auditory cortex and levels of processing in primates. *Audiol Neurootol* 3: 73–85, 1998.
- Kanold PO, Nelken I, Polley DB.** Local versus global scales of organization in auditory cortex. *Trends Neurosci* 37: 502–510, 2014.
- Kikuchi Y, Horwitz B, Mishkin M.** Hierarchical auditory processing directed rostrally along the monkey's supratemporal plane. *J Neurosci* 30: 13021–13030, 2010.
- Kimura A, Donishi T, Sakoda T, Hazama M, Tamai Y.** Auditory thalamic nuclei projections to the temporal cortex in the rat. *Neuroscience* 117: 1003–1016, 2003.
- Knutson B, Burgdorf J, Panksepp J.** Anticipation of play elicits high-frequency ultrasonic vocalizations in young rats. *J Comp Psychol* 112: 65–73, 1998.
- Knutson B, Burgdorf J, Panksepp J.** Ultrasonic vocalizations as indices of affective states in rats. *Psychol Bull* 128: 961–977, 2002.
- Kusmierek P, Rauschecker JP.** Functional specialization of medial auditory belt cortex in the alert rhesus monkey. *J Neurophysiol* 102: 1606–1622, 2009.
- Kvale M, Schreiner C.** Short-term adaptation of auditory receptive fields to dynamic stimuli. *J Neurophysiol* 91: 604–612, 2004.
- Lakatos P, Pincze Z, Fu KM, Javitt DC, Karmos G, Schroeder CE.** Timing of pure tone and noise-evoked responses in macaque auditory cortex. *Neuroreport* 16: 933–937, 2005.
- LeDoux JE, Farb CR, Romanski LM.** Overlapping projections to the amygdala and striatum from auditory processing areas of the thalamus and cortex. *Neurosci Lett* 134: 139–144, 1991.
- Liu RC, Linden JF, Schreiner CE.** Improved cortical entrainment to infant communication calls in mothers compared with virgin mice. *Eur J Neurosci* 23: 3087–3097, 2006.
- Liu RC, Schreiner CE.** Auditory cortical detection and discrimination correlates with communicative significance. *PLoS Biol* 5: e173, 2007.
- Marlin BJ, Mitre M, D'Amour JA, Chao MV, Froemke RC.** Oxytocin enables maternal behaviour by balancing cortical inhibition. *Nature* 520: 499–504, 2015.
- McIntosh TK, Barfield RJ, Geyer LA.** Ultrasonic vocalisations facilitate sexual behaviour of female rats. *Nature* 272: 163–164, 1978.
- Medvedev AV, Kanwal JS.** Local field potentials and spiking activity in the primary auditory cortex in response to social calls. *J Neurophysiol* 92: 52–65, 2004.
- Natan RG, Briguglio JJ, Mwilambwe-Tshilobo L, Jones S, Aizenberg M, Goldberg EM, Geffen MN.** Complementary control of sensory adaptation by two types of cortical interneurons. *eLife* 2015: 10.7554/eLife.09868, 2015.
- Otazu GH, Tai LH, Yang Y, Zador AM.** Engaging in an auditory task suppresses responses in auditory cortex. *Nat Neurosci* 12: 646–654, 2009.
- Parrott RF.** Effect of castration on sexual arousal in the rat, determined from records of post-ejaculatory ultrasonic vocalizations. *Physiol Behav* 16: 689–692, 1976.
- Paxinos G, Watson C.** *The Rat Brain in Stereotaxic Coordinates.* Sydney: Academic, 1986.
- Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E.** Scikit-learn: Machine Learning in Python. *J Mach Learn Res* 12: 2825–2830, 2011.
- Pelleg-Toiba R, Wollberg Z.** Discrimination of communication calls in the squirrel monkey: “call detectors” or “cell ensembles”? *J Basic Clin Physiol Pharmacol* 2: 257–272, 1991.
- Petkov CI, Kayser C, Augath M, Logothetis NK.** Functional imaging reveals numerous fields in the monkey auditory cortex. *PLoS Biol* 4: e215, 2006.
- Pincherli Castellanos TA, Aitoubah J, Molotchnikoff S, Lepore F, Guillemot JP.** Responses of inferior collicular cells to species-specific vocalizations in normal and enucleated rats. *Exp Brain Res* 183: 341–350, 2007.
- Polley DB, Read HL, Storace DA, Merzenich MM.** Multiparametric auditory receptive field organization across five cortical fields in the albino rat. *J Neurophysiol* 97: 3621–3638, 2007.
- Portfors CV.** Types and functions of ultrasonic vocalizations in laboratory rats and mice. *J Am Assoc Lab Anim Sci* 46: 28–34, 2007.
- Profant O, Burianova J, Syka J.** The response properties of neurons in different fields of the auditory cortex in the rat. *Hear Res* 296: 51–59, 2013.
- Rabinowitz NC, Willmore BD, King AJ, Schnupp JW.** Constructing noise-invariant representations of sound in the auditory pathway. *PLoS Biol* 11: e1001710, 2013.
- Rabinowitz NC, Willmore BD, Schnupp JW, King AJ.** Contrast gain control in auditory cortex. *Neuron* 70: 1178–1191, 2011.
- Rauschecker JP, Tian B.** Processing of band-passed noise in the lateral auditory belt cortex of the rhesus monkey. *J Neurophysiol* 91: 2578–2589, 2004.
- Rauschecker JP, Tian B, Hauser M.** Processing of complex sounds in the macaque nonprimary auditory cortex. *Science* 268: 111–114, 1995.
- Recanzone GH, Cohen YE.** Serial and parallel processing in the primate auditory cortex revisited. *Behav Brain Res* 206: 1–7, 2010.
- Roger M, Arnault P.** Anatomical study of the connections of the primary auditory area in the rat. *J Comp Neurol* 287: 339–356, 1989.

- Romanski LM, LeDoux JE.** Information cascade from primary auditory cortex to the amygdala: corticocortical and corticoamygdaloid projections of temporal cortex in the rat. *Cereb Cortex* 3: 515–532, 1993a.
- Romanski LM, LeDoux JE.** Organization of rodent auditory cortex: anterograde transport of PHA-L from MGv to temporal neocortex. *Cereb Cortex* 3: 499–514, 1993b.
- Rust NC, DiCarlo JJ.** Balanced increases in selectivity and tolerance produce constant sparseness along the ventral visual stream. *J Neurosci* 32: 10170–10182, 2012.
- Rust NC, Dicarlo JJ.** Selectivity and tolerance (“invariance”) both increase as visual information propagates from cortical area V4 to IT. *J Neurosci* 30: 12978–12995, 2010.
- Rutkowski RG, Miasnikov AA, Weinberger NM.** Characterisation of multiple physiological fields within the anatomical core of rat auditory cortex. *Hear Res* 181: 116–130, 2003.
- Sales GD.** Ultrasound and mating behavior in rodents with some observations on other behavioral situations. *J Zool* 68: 149–164, 1972.
- Sally S, Kelly J.** Organization of auditory cortex in the albino rat: sound frequency. *J Neurophysiol* 59: 1627–1638, 1988.
- Schneider DM, Woolley SM.** Discrimination of communication vocalizations by single neurons and groups of neurons in the auditory midbrain. *J Neurophysiol* 103: 3248–3265, 2010.
- Schnupp JW, Hall TM, Kokelaar RF, Ahmed B.** Plasticity of temporal pattern codes for vocalization stimuli in primary auditory cortex. *J Neurosci* 26: 4785–4795, 2006.
- Sewell GD.** Ultrasonic communication in rodents. *Nature* 227: 410, 1970.
- Sharpee T, Atencio C, Schreiner C.** Hierarchical representations in the auditory cortex. *Curr Opin Neurobiol* 21: 761–767, 2011.
- Sirotin YB, Costa ME, Laplagne DA.** Rodent ultrasonic vocalizations are bound to active sniffing behavior. *Front Behav Neurosci* 8: 399, 2014.
- Storage DA, Higgins NC, Read HL.** Thalamic label patterns suggest primary and ventral auditory fields are distinct core regions. *J Comp Neurol* 518: 1630–1646, 2010.
- Takahashi N, Kashino M, Hironaka N.** Structure of rat ultrasonic vocalizations and its relevance to behavior. *PLoS One* 5: e14115, 2010.
- Ulanovsky N, Las L, Farkas D, Nelken I.** Multiple time scales of adaptation in auditory cortex neurons. *J Neurosci* 24: 10440–10453, 2004.
- Vapnik V.** *The Nature of Statistical Learning*. Springer Verlag, 2000.
- Wallace MN, Shackleton TM, Anderson LA, Palmer AR.** Representation of the purr call in the guinea pig primary auditory cortex. *Hear Res* 204: 115–126, 2005.
- Wang X, Merzenich MM, Beitel R, Schreiner CE.** Representation of a species-specific vocalization in the primary auditory cortex of the common marmoset: temporal and spectral characteristics. *J Neurophysiol* 74: 2685–2706, 1995.
- Winer JA, Kelly JB, Larue DT.** Neural architecture of the rat medial geniculate body. *Hear Res* 130: 19–41, 1999.
- Winer JA, Schreiner CE.** *The Auditory Cortex*. New York: Springer, 2010.
- Wohr M, Houx B, Schwarting RK, Spruijt B.** Effects of experience and context on 50-kHz vocalizations in rats. *Physiol Behav* 93: 766–776, 2008.
- Woolley S, Fremouw T, Hsu A, Theunissen F.** Tuning for spectro-temporal modulations as a mechanism for auditory discrimination of natural sounds. *Nat Neurosci* 8: 1371–1379, 2005.
- Wright JM, Gourdon JC, Clarke PB.** Identification of multiple call categories within the rich repertoire of adult rat 50-kHz ultrasonic vocalizations: effects of amphetamine and social context. *Psychopharmacology* 211: 1–13, 2010.

