1-1-2015

Encoding of Ultrasonic Communication Signals in Rat Auditory Cortex

Isaac Michael Carruthers
University of Pennsylvania, isaac.carruthers@gmail.com

Follow this and additional works at: http://repository.upenn.edu/edissertations
Part of the Computer Sciences Commons, and the Neuroscience and Neurobiology Commons

Recommended Citation

This paper is posted at ScholarlyCommons. http://repository.upenn.edu/edissertations/1023
For more information, please contact repository@pobox.upenn.edu.
Encoding of Ultrasonic Communication Signals in Rat Auditory Cortex

Abstract
All social animals require a means of communication, and for many species that need is filled by the use of vocalizations. While far less intricate than human speech, many animals employ systems of vocalizations in order to attract mates, convey information about the environment, or to express an emotional state. One such animal is the rat, which communicates via a set of ultra-sonic vocalizations (USVs) in the 50kHz frequency range. These USVs have a conveniently simple structure, making them easy to synthesize and modify. The rat thus provides an excellent model system with which to probe the processing and encoding of such communication signals in the mammalian brain.

In the studies contained within this work we take several novel steps in the investigation of rat vocalizations, and in the study of auditory objects in general. We develop a novel system for parameterizing, purifying, and modifying rat USVs. We model neural responses to USVs, and show that a simple model based on frequency modulation outperforms a more traditional, spectral-based model. We study how neurons in the auditory cortex react to shifts in the statistical structure of USVs, and find evidence that the primary auditory cortex is specialized for the temporal structure of natural vocalizations. We go on to examine the degree to which neural representations of USVs are invariant to small transformations of the USVs, and find evidence that this invariance is greater in the higher brain area SRAF, than in the lower brain area A1. Finally, we develop and implement an experimental system that allows us to probe a rat's perception of a stimulus by examining the rat's behavioral reactions.

Degree Type
Dissertation

Degree Name
Doctor of Philosophy (PhD)

Graduate Group
Physics & Astronomy

First Advisor
Maria N. Geffen

Keywords
Auditory Neuroscience, Computational Neuroscience, Sensory Neuroscience

Subject Categories
Computer Sciences | Neuroscience and Neurobiology

This dissertation is available at ScholarlyCommons: http://repository.upenn.edu/edissertations/1023
ENCODING OF ULTRASONIC COMMUNICATION SIGNALS IN RAT AUDITORY CORTEX

Isaac Michael Carruthers

A DISSERTATION

in

Physics and Astronomy

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2015

Supervisor of Dissertation

______________________________

Maria Neimark Geffen

Assistant Professor of Neuroscience and Otorhinolaryngology

Graduate Group Chairperson

______________________________

Marija Drndic

Professor and Graduate Chair of Physics and Astronomy

Dissertation Committee

Philip Nelson, Professor of Physics and Astronomy
Vijay Balasubramanian, Cathy and Marc Lasry Professor of Physics and Astronomy
Professor Mark Goulian, Professor of Physics and Astronomy
Alison Sweeney, Assistant Professor of Physics and Astronomy
ENCODING OF ULTRASONIC COMMUNICATION SIGNALS IN RAT AUDITORY CORTEX

© 2015

Isaac Michael Carruthers

This work is licensed under

The Creative Commons Attribution-NonCommercial-ShareAlike license, v4.0

To view a copy of this license, visit

http://creativecommons.org/licenses/by-nc-sa/4.0/
DEDICATION

For Maria, without whose unfathomable patience I doubt I would have come this far.
ACKNOWLEDGMENTS

The work contained herein represents the combined effort, advice, guidance, and supervision of so many wonderful scientists that I dare not try to name them all. I owe a great debt of gratitude to all past and present members of the Geffen lab, who were always there to talk and to lend a helping hand; to my professors in the Physics department, who encouraged me to apply myself outside of the boundaries of my own discipline; and of course to my parents, who instilled scientific principles in my young mind before it had fully formed.
ABSTRACT

ENCODING OF ULTRASONIC COMMUNICATION SIGNALS IN RAT AUDITORY CORTEX

Isaac Michael Carruthers
Maria Neimark Geffen

All social animals require a means of communication, and for many species that need is filled by the use of vocalizations. While far less intricate than human speech, many animals employ systems of vocalizations in order to attract mates, convey information about the environment, or to express an emotional state. One such animal is the rat, which communicates via a set of ultra-sonic vocalizations (USVs) in the 50kHz frequency range. These USVs have a conveniently simple structure, making them easy to synthesize and modify. The rat thus provides an excellent model system with which to probe the processing and encoding of such communication signals in the mammalian brain.

In the studies contained within this work we take several novel steps in the investigation of rat vocalizations, and in the study of auditory objects in general. We develop a novel system for parameterizing, purifying, and modifying rat USVs. We model neural responses to USVs, and show that a simple model based on frequency modulation outperforms a more traditional, spectral-based model. We study how neurons in the auditory cortex react to shifts in the statistical structure of USVs, and find evidence that the primary auditory cortex is specialized for the temporal structure of natural vocalizations. We go on to examine the degree to which neural representations of USVs are invariant to small transformations of the USVs, and find evidence that this invariance is greater in the higher brain area SRAF, than in the lower brain area A1. Finally, we develop and implement an experimental system that allows us to probe a rat’s perception of a stimulus by examining the rat’s behavioral reactions.
TABLE OF CONTENTS

DEDICATION .................................................................................................................. III

ACKNOWLEDGMENTS ................................................................................................. IV

ABSTRACT ....................................................................................................................... V

TABLE OF CONTENTS ................................................................................................. VI

INTRODUCTION ............................................................................................................ 1

1: MODELS OF NEURAL RESPONSES TO VOCALIZATIONS IN A1 ..................... 4

Overview ...................................................................................................................... 4

Methods ....................................................................................................................... 5
  Animals ....................................................................................................................... 5
  Surgery ....................................................................................................................... 5
  Original vocalizations .............................................................................................. 7
  Long vocalization sequence ................................................................................... 9
  Neural recordings .................................................................................................... 11
  Quantification of the neural response strength ...................................................... 12
  Tone pip response measurement .......................................................................... 12
  Frequency-modulation sweep responses ............................................................... 13
  Reduced parameter generalized linear model ..................................................... 15
  Spectrogram-based LNM ...................................................................................... 17
  Statistical tests ....................................................................................................... 17

Results ......................................................................................................................... 18
  A1 neurons exhibit reliable, selective responses to ultra-sonic vocalizations .......... 18
  The differential response of A1 neurons to the eight selected ultra-sonic vocalizations not correlated with their best frequency .................................................. 21
  A1 neurons reliably follow a sequence of vocalizations ......................................... 24
  A1 responses are accurately predicted by a generalized linear model (GLM) ........ 27
  GLM prediction depends on the tuning of A1 units to ultra-sonic frequencies ...... 32
  A1 multi-units respond stronger to original as compared to transformed USVs ....... 34
  GLM predicts A1 neuronal responses less accurately to transformed than to original vocalizations .............................................................................................................. 39

Discussion .................................................................................................................... 40
  Preference of A1 responses for temporal structure of original USVs ..................... 41
  Similar computation underlying responses to reversed vocalizations ................. 43
  Response strength of A1 neurons to USVs ............................................................ 43
2: EMERGENCE OF INVARIANT REPRESENTATIONS OF VOCALIZATIONS ............................. 47

Overview .................................................................................................................................. 47

Introduction ............................................................................................................................... 47

Methods .................................................................................................................................... 51
   Animals. ................................................................................................................................... 51
   Stimuli. .................................................................................................................................... 51
   Stimulus Transformations ......................................................................................................... 52
   Microdrive implantation ........................................................................................................... 52
   Stimulus presentation ............................................................................................................... 53
   Electrophysiological recording ................................................................................................. 54
   Unit Selection and Firing-rate Matching .................................................................................. 54
   Response Sparseness ............................................................................................................... 55
   Population Response Vector .................................................................................................... 55
   Linear Support Vector Machine (SVM) Classifier ....................................................................... 56
   Classification Procedure .......................................................................................................... 56
   Bootstrapping .......................................................................................................................... 56
   Mode of Classification .............................................................................................................. 57
   Generalization .......................................................................................................................... 57
   Within-transformation Performance .......................................................................................... 57
   Generalization Penalty .............................................................................................................. 58

Results ....................................................................................................................................... 58

Discussion ................................................................................................................................... 71

3: BEHAVIORAL MEASUREMENTS OF VOCALIZATION DETECTION .......................... 75

Overview ..................................................................................................................................... 75

System Design ............................................................................................................................. 76

Experimental Protocol ................................................................................................................ 81

Results ....................................................................................................................................... 87

BIBLIOGRAPHY ......................................................................................................................... 90
INTRODUCTION

The mammalian auditory system is responsible for the somewhat daunting task of transforming spectrally complex pressure waveforms into useful information about the physical world. At the level of the cochlea and the auditory nerve, neural activity primarily conveys raw information about the spectral content of incoming sound. However, by the time auditory neural signals are integrated into decision-making processes, they contain information about complex object relationships. Such information may include whether an incoming sound might have been made by a hard object striking hollow wood, or whether the sound was a spoken word. Such object representations must be invariant to common distortions that occur in natural environments (Sharpee et al. 2011); a word must still sounds like a word when spoken by a different speaker, or when the wind is blowing in the background. Much of the intermediate processing that transforms the feature-based representation of the periphery to the invariant, object-based representation of the higher brain areas is poorly understood.

Rats provide us with an excellent model system with which to probe the formation of auditory objects. Rats communicate via a diverse, but spectrally simple set of vocalizations in the 50kHz frequency range (Knutson et al. 2002; Portfors 2007; Sewell 1970; Takahashi et al. 2010). Male rats emit these ultra-sonic vocalizations (USVs) in reaction to many positive stimuli, including those associated with food and sex (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; Wohr et al. 2008). While immediate early gene expression has been shown to be elevated in A1 following exposure to USVs (Sadananda et al. 2008), the neural
correlates of responses to USVs in rats have previously been identified only in the perirhinal cortex (Allen et al. 2007) and the amygdala (Parsana et al. 2012). Understanding how neurons in A1 encode vocalizations is essential for comprehending the function of areas that receive direct and indirect input from A1, and how perceptual correlates of vocalizations are formed in the downstream areas (Doupe and Kuhl 1999). In chapter one of this manuscript, we characterize the responses of A1 neurons to USVs, and develop models of their feature selectivity, thus contributing to our understanding of how these behaviorally important stimuli are processed.

It has been proposed that the auditory system (in addition to the visual system and others) builds up an object-based representation of stimuli in a hierarchical fashion. Within this theory, each stage of neural processing performs transformations on the neural representation of the auditory scene, incrementally building object-invariant responses. The auditory cortex (AC) in particular has been shown to be an essential stage in the encoding of auditory objects (Aizenberg 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). For example, although neurons in input layers of the primary auditory cortex (A1) preferentially respond to specific features of acoustic stimuli, neurons in the output layers tend to be more selective to combinations of features (Atencio et al. 2009; Sharpee et al. 2011). Similar studies in the visual pathway have also supported this hierarchical theory of encoding (DiCarlo and Cox 2007; Rust and DiCarlo 2012; 2010). In chapter one of this manuscript, we examine the combinations of auditory features that are represented in the primary auditory cortex. In chapter two, we go on to investigate whether the progressive development of object-invariance observed in A1 continues as we move on to the supra-rhinal auditory field (SRAF, a non-primary auditory area).
The study of auditory cortical responses, described thus far, would be incomplete without an examination of how the activity in this area correlates with behavior. The ultimate goal of all neural processing is to allow the organism to better interact with its environment, and so if a neural computation is not used to influence the organism’s behavior then it is in most senses irrelevant. It has been shown, for instance, that animals do not always perform tasks with the level of accuracy that would be predicted from recording their neural activity (Carney et al. 2014). It is therefore necessary to compare our neural studies with behavioral results, in order to fully understand the animals’ ability to process auditory stimuli. To this end, in chapter three of this manuscript, we detail the development of a behavioral training system that will allow us to compare our neural recordings to our animals’ actual ability to perform tasks based on auditory stimuli.
1: MODELS OF NEURAL RESPONSES TO VOCALIZATIONS IN A1

Overview

One of the central tasks of the mammalian auditory system is to represent information about acoustic communicative signals, such as vocalizations. However, the neuronal computations underlying vocalization encoding in the central auditory system are poorly understood. To learn how the rat auditory cortex encodes information about conspecific vocalizations, we presented a library of natural and temporally transformed ultrasonic vocalizations (USVs) to awake rats while recording neural activity in the primary auditory cortex (A1) with chronically implanted multielectrode probes. Many neurons reliably and selectively responded to USVs. The response strength to USVs correlated strongly with the response strength to frequency-modulated (FM) sweeps and the FM rate tuning index, suggesting that related mechanisms generate responses to USVs as to FM sweeps. The response strength further correlated with the neuron’s best frequency, with the strongest responses produced by neurons whose best frequency was in the ultrasonic frequency range. For responses of each neuron to each stimulus group, we fitted a novel predictive model: a reduced generalized linear model (GLM) that takes the frequency modulation and single-tone amplitude as the only two input parameters. The GLM accurately predicted neuronal responses to previously unheard USVs, and its prediction accuracy was higher than that of an analogous spectrogram-based linear-nonlinear model. The response strength of neurons and the model prediction accuracy were higher for original, rather than temporally transformed, vocalizations. These results indicate that A1 processes original USVs differentially than transformed USVs, indicating preference for temporal statistics of the original vocalizations.
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 rats. Rats were housed in a temperature and humidity-controlled vivarium on a reversed 24 hour light-dark cycle with food and water provided ad libiditum.

Surgery

Sprague Dawley or Long Evans adult male rats (N = 6, 12-16 weeks) were anesthetized with an intra-peritoneal injection of a mixture of ketamine (60 mg per kg of body weight) and dexmedetomidine (0.25 mg per kg). Buprenorphine (0.1 mg/kg) was administered as an operative analgesic with Ketoprofen (5 mg/kg) as post-operative analgesic. Rats were implanted with chronic custom-built multi-tetrode microdrives as previously described (Otazu et al. 2009). The animal’s head was secured in a stereotactic frame (David Kopf Instruments). Following the recession of the temporal muscle, a craniotomy and durotomy were performed over the location of A1. A microdrive housed eight tetrodes, of which two were used for reference and six for signal channels. Each tetrode consisted of four polyimide-coated Nichrome wires (Kenthal-PalmCoast, wire diameter of 12 microns) twisted together, and was controlled independently with a turn of a screw. Two screws (one reference and one ground) were inserted in the skull at locations distal from A1. The tetrodes were positioned 4.0-6.5 mm posterior to bregma and 6.0 mm left of the midline, covered with agar solution (3.5%), and the microdrive was secured to the skull with dental acrylic (Metabond) and dental cement. The location of the electrodes was verified based on the stereotaxic coordinates, the electrode position in relation
to brain surface blood vessels, and through histological reconstruction of the electrode tracks (Figure 1A). During the recording, the microdrive was connected via a custom-built interface board to a headstage (Neuralynx Inc.). The electrodes were gradually advanced below the brain surface in daily increments of 40-50 micrometers. The location was also confirmed by identifying the frequency tuning curve of the recorded units (Figure 1B). The recorded units' best frequency (frequency of the tone which elicited the highest firing rate (see below)) spanned the range of rat hearing (Figure 1C) and was consistent with previous studies (Sally and Kelly 1988).

Figure 1. Recording location and tuning of recorded units. A. Trace of the tetrode in a Nissl-stained, fixed coronal slice of the brain. The tetrode terminated in layer 5 of A1 (arrow). B. Tuning curve of a representative unit. The color represents the response strength to a tone pip at a specific frequency (x-axis) and loudness (y-axis). C. Distribution of the best frequency and tuning bandwidth of all recorded units that had significant response to the tuning curve stimulus. D. The responses of a representative, FM sweep rate-tuned unit to FM sweeps. Left:
timecourse of the firing rate of the unit in response to FM sweeps at different rates. Red line indicated stimulus onset, and red dots indicate stimulus offset. Right: Response strength of the unit to FM sweeps at different rates. E.

The responses of a representative, FM direction-tuned unit to FM sweeps. Axes same as D.

Original vocalizations

The original vocalizations were extracted from a recording provided by Diego A. Laplagne (Rockefeller University). The recordings were collected when two adult male rats, housed in isolation, were placed in a single cage together for 2 hours. Vocalizations were recorded using a free-field ultra-sonic microphone (AvisoBioacoustics, CM15, sensitivity 50 mV/Pa, frequency range: 10 - 200 kHz, input-referred self-noise level 18 dB).

From the continuous recording, vocalizations were extracted for further analysis, separately for each rat. The recorded sound wave was transformed into a spectrogram using the multi-tapered spectrogram transform (Chronux toolbox, (Bokil et al. 2010)), the entropy of the signal across all spectral channels was computed, and subjected to a threshold. The onset of the vocalizations was taken as the time at which the threshold was reached. The threshold was manually adjusted to capture all vocalizations that were visually observed as distinct in the spectrogram of the signal, after which the analysis was fully automatic. The minimum inter-vocalization separation for detection was set to 40 ms, so the onset of each vocalization was identified by a threshold crossing that was at least 40 ms after the previous time the spectral entropy exceeded threshold. For initial response characterizations, eight vocalizations were isolated at random from the long recording. 350 additional vocalizations were isolated for subsequent response characterization.
A noiseless version of the vocalizations was constructed to ensure that the neural responses were due to vocalizations and not due to an interaction with background noise (Bar-Yosef and Nelken 2007). To generate the noiseless stimuli, using an automated procedure, we isolated the dominant frequency and amplitude for each noisy vocalization (Figure 2). The noiseless signal was constructed 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. In each 1.0 ms bin, the values of $f(t)$ and $a(t)$ were extracted from a multi-tapered spectrogram of the vocalizations, generated using a 2.0 ms window in steps of 0.25 ms, by convolving each temporal slice of the spectrogram with a ridge-detecting filter (Figure 2B). They were then resampled back up to our system playback frequency (400 kHz) using a shape-preserving piecewise cubic interpolation (Figure 2C). The noiseless signal was generated as

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

(Figure 2D). The power spectrum of the noiseless vocalization matched that of the recorded in the USV range, whereas the noise power sidebands were removed (Figure 2E). Eight noiseless vocalizations were constructed from the originally recorded vocalizations (Figure 4A).
Long vocalization sequence

Three hundred and fifty vocalizations were extracted from the original vocalization recording, and noiseless versions of these vocalizations were constructed as described above. Next, the vocalizations were concatenated into a long string with 50 ms inter-vocalization separation (Figure 3A, 6A, 8A). The 50 ms inter-vocalization separation was chosen to match the mean natural vocalization rate of 10 Hz (unpublished observations, Diego A. Laplagne). The temporally dilated vocalizations (Figure 3B) were generated as

\[ x(t) = a(0.67t)\sin \left( 2\pi \int_{0}^{0.67t} f(\tau) d\tau \right) \]

and temporally compressed vocalizations (Figure 3C)

\[ x(t) = a(1.5t)\sin \left( 2\pi \int_{0}^{1.5t} f(\tau) d\tau \right) \]

were generated as. To generate the reverse vocalization sequence, the original calls were reversed in time,

\[ x(T - t) = a(t)\sin \left( 2\pi \int_{0}^{t} f(\tau) d\tau \right) \], and
concatenated in the opposite order to form a sequence (Figure 3D). The temporal and spectral modulation power spectrum was computed as the Fourier transform of the auto-correlation matrix of the spectrogram of the full stimulus (Singh and Theunissen 2003). The range of the temporal and frequency modulations of the stimuli differed for temporally compressed and dilated stimuli as compared to the original and reversed (Figure 3, right panels).

Figure 3. Spectro-temporal content of original and transformed vocalizations. A. Left: Spectrogram of a sequence of 4 original vocalizations of the 350 vocalization sequence. Each vocalization is represented as a continuous amplitude and frequency-modulated tone. Right: temporal and frequency modulation spectrum of the 350
vocalizations. B. Spectrogram and modulation spectrum of temporally compressed (accelerated, x1.5),
vocalizations. C. Spectrogram and modulation spectrum of temporally dilated (slowed down, x0.67) vocalizations.
D. Spectrogram and modulation spectrum of reversed vocalizations.

**Neural recordings**

Neural signals were acquired daily from 24 chronically implanted electrodes in awake, freely-moving rodents using a Neuralynx Cheetah system. The neuronal signal was filtered between 0.6 kHz and 6.0 kHz, digitized and recorded at 32 kHz rate. Spikes were clustered into single-unit and multi-unit clusters using either Neuralynx Spike Sort 3D or Plexon Off-line Spike Sorter software. We used a stringent set of criteria to isolate single units from multi-unit clusters (Bizley et al. 2010; Brasselet et al. 2012; Otazu et al. 2009). Single-unit clusters contained < 0.1% of spikes within 1.0 ms inter-spike interval, and the spike waveforms had to form a visually identifiable distinct cluster in a projection onto a three-dimensional subspace.

The acoustical stimulus was delivered via a magnetic speaker (MF-1, Tucker-Davis Technologies) positioned above the recording chamber. The speaker output was calibrated using Brue and Kjaer 1/4 inch free-field microphone type 4939, which was placed at the location that would normally be occupied by the animal’s ear, by presenting an recording the speaker output of repeated white noise bursts and tone pips between 400 and 80000 Hz. From these measurements, the speaker transfer function and it inverse were computed. The input to the microphone was adjusted using the inverse of the transfer function such that the speaker output 70 dB tones within 3dB between 400 and 80000 Hz. Spectral and temporal distortion products were measured in response to tone pips between 1 and 80 kHz, and were found to be > 50 dB below the SPL of the fundamental. All stimuli were presented at 400 kHz sampling rate,
using custom-built software based on a commercially available data acquisition toolbox (Mathworks, Inc.), and a high-speed data acquisition card (National Instruments, Inc.).

Quantification of the neural response strength

To compute the strength of neuronal responses to the individual USVs, the responses of neurons to 50-200 repeats of the eight USVs were recorded, and binned in 10 ms bins. The baseline was taken 0.5 - 1.0 s after the vocalization onset. Each response was represented as a vector consisting of spike counts in 10 ms bins between 10 and 120 ms post-stimulus onset. The minimum firing rate was set to 0.1 Hz during the response and at baseline. The response strength was calculated as the Mahalanobis distance between the response and the baseline. The Mahalanobis distance $D_M$ of a vector $\mathbf{x}$ assumed to be from a distribution with mean $\mu$ and covariance matrix $S$ is computed as $D_M(\mathbf{x}) = \sqrt{(\mathbf{x} - \mu)^T S^{-1} (\mathbf{x} - \mu)}$. Responses were considered significant if this measure, normalized by the square root of the stimulus repeat number, exceeded 3. The response-eliciting USV fraction, RUSV, was computed for each neuron as the number of USVs eliciting significant responses divided by the number of USVs presented. The response selectivity index, RS, was computed as the maximum response strength to a USV divided by the mean response strength to all USVs presented.

Tone pip response measurement

The firing rate of neurons was recorded in response to randomly interleaved 50 ms long tone pips, with 250 ms inter-tone interval. For a subset of neurons ($N = 147$), amplitude- and frequency- tuning curves were collected: tone pips at 100 frequencies spaced uniformly in log-frequency space between 0.4 and 80 kHz were presented at 10 sound pressure levels each,
uniformly distributed between 10 and 80 dB (relative to reference pressure of 20 mPa). The best frequency was computed as the frequency of the tone that evoked the maximum response strength averaged over SPL of 40 and 80 dB (Figure 1B). For another subset of units (N = 424), tone pips were presented at a single sound pressure level of 50 dB (relative to reference pressure of 20 mPa) with 100 frequencies spaced uniformly in log-frequency space between 0.4 and 80 kHz. The response strength, which combined onset and offset responses, was computed as the mean firing rate of neurons during 0 to 80 ms after tone onset. The best frequency was computed as the frequency of the tone that evoked the maximum response strength (Brown and Harrison 2009). The tuning bandwidth was computed at 10% of the maximum of the peak, fitted to a Gaussian. The peak was considered significant if the maximum firing rate exceeded by 3 standard deviations the firing rate in response to frequencies outside the peak. The overlap between the spectral response profile and the power spectrum of vocalizations was computed as the dot product of the power spectrum of the USV waveform and the response strength of the neuron at each frequency (extrapolated to the frequencies of the USV power spectrum), normalized by the sum of the power spectrum of the USV across all frequencies.

**Frequency-modulation sweep responses**

The spiking responses of neurons were recorded in response to randomly interleaved frequency-modulated (FM) sweeps. The sweeps were presented at 500 ms interval between the end and the beginning of successive sweeps. The sweeps were composed as a tone, whose frequency was swept logarithmically between 1kHz and 80kHz. The sweep was padded at each end with 100 ms pure tone at the start or end frequency (1kHz or 80kHz, depending on the sweep direction). The sweeps were presented at 22 rates, log-uniformly distributed between
+64 and -64 Octaves/s. The firing rate for each FM sweep rate was computed by binning the spikes in 10 ms bins, and smoothing them with a 3-bin Gaussian envelope. The response strength to each FM sweep rate, $R_i$, was computed as the normalized mean baseline-subtracted firing rate in bins during the sweep presentation, during which the firing rate exceeded the 95% confidence limit of the baseline firing rate. The firing rate was normalized by the standard deviation of the baseline firing rate to facilitate comparison across units. A small offset term was added to the denominator to prevent division by 0. The 95% confidence limit was computed over the baseline firing rate, assuming the baseline fluctuated as a Gaussian with the standard deviation computed over all trials. The FM rate tuning index, $IFM$, was computed over $n$ sweep speeds in which the firing rate exceeded the 95% confidence limit (Atencio et al. 2007):

$$IFM_{tuning} = \frac{n}{n-1} \left( 1 - \frac{\langle R_i \rangle}{\max(R_i)} \right)$$

Brackets denote the average over all sweep rates.

The FM directionality index, $ID$, was computed over responses to sweeps where $R_i^+$ denotes the response strength to an up sweep at rate i, and $R_i^-$ denotes the response strength to a down sweep at rate i (Atencio et al. 2007; Nelken and Versnel 2000; Shamma et al. 1993):

$$ID = \frac{\langle R_i^+ \rangle - \langle R_i^- \rangle}{\langle R_i^+ \rangle + \langle R_i^- \rangle}$$

Brackets denote the average over all up or down sweep rates.
**Reduced parameter generalized linear model**

A generalized linear model (GLM) provided an ideal framework for constructing a predictive model of neuronal responses to the stimulus (Calabrese et al. 2011a; Pillow et al. 2011; Pillow et al. 2008), because of the non-Gaussian statistics of the stimulus. The output of the model is the activity of an individual neuron; while the input to the model is the stimulus, as represented by its envelope $a(t)$ and differential frequency $\omega'(t)$ (Figure 8B). In the model each of two input parameters is convolved with a linear kernel, after which the outputs of the filters are summed (Figure 8C). This sum then undergoes a rectifying non-linearity to approximate the spiking transformation (Figure 8D). The output of this function is entered into a Poisson generator to predict individual neuronal firing. The model prediction is calculated as:

$$r_i(t) = P \left( f \left( b + \int_{\tau=0}^{T} a(t - \tau)k_i^a(\tau) + \omega'(t - \tau)k_i^\omega(\tau) d\tau \right) \right)$$

where $r_i(t)$ is the response of neuron $i$ at time $t$, $k_i$ is the linear kernel of the neuron with respect to the stimulus, $T$ is the length of the kernels, $b$ is the baseline log-firing rate, $f(x)$ is the instantaneous non-linear function (here taken as an exponential), and $P$ is a Poisson generator.

To fit GLM, we used the maximum likelihood optimization approach (Pillow et al. 2008). With this approach, the model parameters are determined such that they maximize the likelihood of the recorded spike trains given the prediction of the model. We approximated the log-likelihood of the spike train as $L = \sum_{t_i} \ln r_i(t_i) - \int r_i(t)dt$ where $t_i$ are the spike times. We then calculated the gradient of the log-likelihood with respect to the model parameters, and used standard iterative optimization algorithms to find the optimal model, maximizing the
likelihood of the spiking response over 100 trials given the model's prediction. To be included in the analysis, a unit had to have a mean discharge rate of at least 0.1 Hz during the stimulus. At each iteration the algorithm computed the log-likelihood and the gradient of the log-likelihood with respect to the model parameters, and incremented the model's parameters along the steepest gradient. Increments were determined using the built-in Matlab optimization procedure fminunc (Mathworks, Inc.).

When maximizing $L$ above, we also included an L2 regularization term: $-\gamma \int_0^T k_t^\alpha(t)^2 + k_t^\beta(t)^2 \, dt$, which served to reduce over-fitting noise in the model. Due to computational constraints, the regularization coefficient $\gamma$ was determined empirically by selecting the value that resulted in the model with the highest predictive accuracy in the case of a few exemplar cells. The accuracy of prediction of the model was computed as the coefficient of correlation between the model prediction and recorded responses to a novel stimulus.

For each neuron, spikes were binned in 1 ms bins and smoothed with a Gaussian window of 1 ms width. Because recent studies indicate that neurons in A1 carry information at less than 2 ms precision (Kayser et al. 2010), the smallest bin size (1 ms) was used for the analysis. Firing rate was computed for each trial, and as an average across trials. The response strength to the USVs, $R_i$, was computed as the normalized mean baseline-subtracted firing rate in bins during the USV presentation, during which the firing rate exceeded the 95% confidence limit of the baseline firing rate. The firing rate was normalized by the standard deviation of the baseline firing rate to facilitate comparison across units. A small offset term was added to the denominator to prevent division by 0. The 95% confidence limit was computed over the baseline
firing rate, assuming the baseline fluctuated as a Gaussian with the standard deviation computed over all trials. The significance of the response of the neuron was assayed by the signal-to-noise ratio, defined as the ratio of the standard deviation of the firing rate averaged over the vocalization sequence (corresponding to the square root of the power of the signal), divided by the standard error of the mean firing rate across trials (noise). A neuron was considered to respond significantly if the signal-to-noise ratio was higher than 2. The model was fitted on 200 vocalizations and the predictive accuracy was computed as the mean over the remaining 150 vocalizations.

Spectrogram-based LNM

To analyze the improvement in the model fit due to the low-dimensional parameterization of the stimulus, we also fitted a linear – non-linear model (LNM) computed using standard reverse correlation technique (Baccus and Meister 2002; Escabi and Read 2003; Geffen et al. 2007; Theunissen et al. 2001), using a spectrogram as an input (Figure 8E). The filter was computed by normalizing the convolution of the response and the stimulus by the stimulus auto-correlation matrix (Figure 8F), and the instantaneous non-linearity (Figure 8G) was computed directly from firing rate versus linear prediction plot (Baccus and Meister 2002; Geffen et al. 2007).

Statistical tests

The correlation coefficient (r) was computed as Pearson's correlation coefficient using standard Matlab routines. Student t-test and multi-variate analysis of variance (MANOVA) were conducted on either paired or unpaired samples (as indicated in text) using standard Matlab
routines. Repeated measure analysis of variance (ANOVA) was performed in SPSS Statistics (IBM, Inc.). Bonferroni multiple comparisons correction was used whenever appropriate.

Results

We measured and analyzed the responses of neurons in A1 to ultra-sonic vocalizations emitted by male rats in a social context. We found that A1 neurons exhibited significant responses to USVs, typically selective for a subset of USVs. The temporal dynamics of A1 responses to a long sequence of USVs presented at the ethological rate were accurately predicted by an integrative model that took the amplitude and frequency modulation as the input (GLM, see Methods). The response strength, as well as the prediction accuracy of the model for each neuron correlated with its FM rate tuning index and best frequency. A1 neurons' response strength and the model prediction accuracy were highest for the original, as compared to temporally transformed USVs, indicating a preference for the ethologically relevant parameters of USVs in the neuronal circuitry that underlies A1 responses.

A1 neurons exhibit reliable, selective responses to ultra-sonic vocalizations

Little is known about how rat USVs are encoded in A1. To assay whether A1 neurons responded to the USVs, we first presented eight vocalizations drawn at random from recordings (Figure 4A), to awake, freely moving rats, and recorded neuronal responses in the primary auditory cortex, using the chronically implanted multi-tetrode microdrives. The units were localized to A1 (Figure 1A), exhibited frequency tuning curves typical of A1 neurons (Figure 1B), and were distributed uniformly in their spectral tuning properties over the rat hearing range (Figure 1C). Neurons exhibited reliable responses to 100 repeats of the stimuli (Figure 4) as
quantified by the signal-to-noise ratio, the standard deviation of the average firing rate divided by the average standard deviation of the firing rate across trials (in 10ms bins), which averaged 2.8 for single units (N = 84) and 2.7 for all units (N = 211). A measure of response strength was used to characterize the neural behavior driven by each of the USVs (see Methods, Figure 5A, B). Figures 4 and 5A, B provide examples that depict the response pattern and normalized response strength in two representative units. Of all units recorded, 27% were responsive (normalized response strength > 3) to at least one vocalization. The results were similar for the single units only: 24% were significantly responsive to at least one vocalization. Unit 1 (Figure 4B, 4C, 5A) exhibited significant responses to vocalizations 2, 4, 5, 6, 7, and 8. Unit 2 (Figure 4D, 4E, 5B) exhibited significant responses to another subset of vocalizations (3, 4, 8). On average, neurons responsive to at least one USV were responsive to 1-2 out of 8 (20%) USVs for multi-units and 2-3 out of 8 (27%) USVs for single units (Figure 5C). These data demonstrate that many A1 neurons significantly responded to the USVs, and that their responses were significant for only a subset of vocalizations.
Figure 4. Responses of two representative neurons to eight vocalizations. A. In each subplot: Top: Waveform of vocalizations. Bottom, left: Spectrogram of the first 100 ms of each vocalization. Bottom, Right: Normalized power spectrum of each vocalization. B. Raster plots of responses of unit 1. C. PSTH of responses of unit 1, binned in 3 ms time bins. D. Raster plots of responses of unit 2, same as B. E. PSTH of responses of unit 2, same as C.

To verify that the denoising procedure did not lead to activation of additional populations of neurons, we presented the original recordings of the USVs (containing background noise) alongside with their denoised versions. The response strength of units was greater for vocalizations containing background noise than to noiseless vocalizations: of all units
recorded in response to noisy vocalizations (N = 395 all units, N = 169 single units), 34% of all units (33% single units) were responsive to at least one noisy vocalization, suggesting that the background increased the responsiveness of neurons to the USVs. The response strength to vocalizations in the presence and absence of noise was significantly correlated over the population of neurons (r = 0.31, p < 1e-38, N = 210, all units; r = 0.39, p < 1e-24, N = 82 single units). These findings are consistent with previous reports on changes of A1 responses to vocalizations upon addition of background (Bar-Yosef and Nelken 2007). Since we were interested in analyzing the responses to USVs without the background, we used denoised vocalizations as stimuli for the remainder of the study.

The differential response of A1 neurons to the eight selected ultra-sonic vocalizations not correlated with their best frequency

Tuning to a specific spectral band is an important response property of A1 neurons (Bizley et al. 2005; Brugge and Merzenich 1973; David et al. 2009; Ehret and Schreiner 1997; Recanzone et al. 1999; Schreiner 1992). We assayed whether the best frequency of A1 units, as determined from responses to tone pips of various frequencies, correlated with the response strength of neurons to ultra-sonic vocalizations (N = 181 all units, N = 82 single units). The proportion of vocalizations to which the unit responded significantly (response-eliciting USV fraction, RUSV) did not correlate significantly with the best frequency of the neurons (Figure 5D; the correlation coefficient was not significant (p > 0.05) for either all units combined or for single units alone). We further computed the overlap between the spectral response profile of the neuron and the power spectrum of each USV, and compared it to RUSV. The spectral response profile was determined as the response strength to tone pips presented at different
frequencies. Across all vocalizations, for neurons that were significantly driven by at least one USV (N = 47 all units, N = 17 single units), RUSV was not significantly correlated with the degree of overlap of the USV power spectrum with the spectral response profile (Figure 5E; the correlation coefficient was not significant (p > 0.05) for either all units combined or for single units alone). The response selectivity index, RS, estimated as the maximum response strength to a USV divided by the mean response strength over all USVs. S also did not exhibit significant correlation (data not shown) with either the best frequency (the correlation coefficient was not significant (p > 0.05) for either all units combined or for single units alone) or the spectral overlap between the USV power spectrum and the spectral response profile of the units (the correlation coefficient was not significant (p > 0.05) for either all units combined or for single units alone).
Figure 5. Response strength of recorded units to the vocalizations. A. Response strength of unit 1 to the vocalizations. B. Response strength of unit 2 to the vocalizations. C. Histogram of the response-eliciting fraction of USVs ($R_{USV}$) for each recorded unit. Red bar - mean of the vocalization selectivity. D. $R_{USV}$ for each unit versus its best frequency. E. $R_{USV}$ of each unit versus the normalized overlap between its spectral response profile and the power spectrum of USVs.

The eight USVs were similar in their power spectrum, with the dominant frequency within half an octave from each other (Figure 4A). The differences among the USVs stemmed primarily from the changes in the temporal structure of the dominant frequency component. Thus, differences in the discharge patterns of those units responsive to the USVs likely were driven by the modulation of the amplitude and frequency modulation in time, rather than the power spectrum.
A1 neurons reliably follow a sequence of vocalizations

To examine the responses of A1 neurons to temporally dynamic USVs, we composed a stimulus sequence, consisting of 350 vocalizations (Figure 3A, 6A, 8A). Figure 6 depicts the responses of two representative units to the vocalization sequence. Both these units exhibited significant responses to the stimulus. The activity of many recorded units (all units N = 397, single units N = 172) was significantly modulated by the stimulus (SNR > 2, 42 % all units, 46% single units).

We compared the response strength of A1 neurons to the long USV sequence with the tuning properties for simpler stimuli, including tone pips and frequency-modulated sweeps.
Units exhibited significant correlation between their best frequency and response strength to USVs, such that units with best frequency in the ultra-sonic range were most responsive to USVs ($r = 0.37, p < 1e-10, N = 284$ all units, $r = 0.27, p < 1e-4, N = 125$ single units, Figure 7A). The normalized spectral overlap between the spectral response profile of the unit and the power spectrum of the USVs was also significantly correlated with response strength to the USVs ($r = 0.34, p < 1e-8$, all units; $r = 0.28, p < 0.005$, single units; Figure 7B). Furthermore, units exhibited high correlation in their response strength to USVs and FM sweeps ($r = 0.63, p < 1e-8, N = 74$ units, $r = 0.41, p < 0.05$ single units, $N = 32$, Figure 7C). The correlation between the FM rate tuning index and USV response strength was weaker across all units ($r = 0.27, p < 0.05$, Figure 7D) and not significant for single units only ($p > 0.05$). This correlation is consistent with the observation that distinct USVs are restricted to a subset of FM rates. No significant correlation was observed between the FM directionality index and USV response strength ($p > 0.05$, Figure 7E), as expected from the symmetric distribution of USV spectral modulation in positive and negative direction (Figure 3, 8B).
Figure 7. Correlation between the response strength to USVs and neuronal tuning properties. A. Response strength to USVs versus best frequency for all units. Each dot represents a single unit. B. Response strength to USVs versus the normalized overlap between its spectral response profile and the power spectrum of USVs. C. Response strength to USVs versus response strength to FM sweeps. D. Response strength to USVs versus response strength to FM rate tuning index. E. Response strength to USVs versus FM directionality index.

A1 responses are accurately predicted by a generalized linear model (GLM)

The computation that underlies generation of responses in A1 to vocalizations remains largely unknown. To characterize the computation by which A1 neurons produce responses to USVs, we next fitted two versions of linear-non-linear model to the responses of each neuron.
during the first 200 vocalizations, and then measured the prediction for the responses to the remaining 150 vocalizations (Figure 8).

While many methods of fitting a linear-non-linear model to neuronal responses exist, the advantage of using the generalized linear model (reduced GLM) for these data is that the probability distribution of the input signal is not required to be Gaussian for the model to converge (Calabrese et al. 2011b; Paninski et al. 2007). We developed a novel version of the GLM, which was based on a low-dimensional representation of the stimulus, including the frequency modulation and the amplitude as functions of time. We compared the prediction accuracy of the GLM to the predictions of a standard linear-non-linear model (STRF LNM) based on the spectrogram of the stimulus. We used standard methods of reverse correlation to fit the spectro-temporal receptive field (STRF) (Escabi and Read 2003; Theunissen et al. 2001) and the instantaneous non-linearity, which was not constrained in its shape (Geffen et al. 2009).

We first present the results of the fit for a representative neuron, and then give the statistics over all recorded units. For the reduced GLM, the stimulus was represented by two parameters: the amplitude and the frequency modulation (Figure 8B). For the STRF LNM, the stimulus was represented as a spectrogram (Figure 8E). Next, the linear filters were fitted under both models. For the sample unit in Figure 8, both the amplitude and frequency modulation (Figure 8C) filters follow an On-type time course (Geffen et al. 2007; Kuffler 1953), in which the peak of the filter is positive. The non-linearity is fitted accurately by an exponential (Figure 8D). In the STRF-based model, the STRF also exhibits a delayed positive peak at about 60kHz, but there is also a negative sideband present around 68 kHz (Figure 8F). The prediction of the GLM
to novel USVs is remarkably accurate (Figure 8H, the fit is in red, the measured firing rate is in black), and more accurate than that of the STRF-based LNM (Figure 8H, green).
Figure 8. Predictive models of A1 responses to vocalizations. A. Sound wave of the stimulus. B. Under GLM, the stimulus is represented as a two-dimensional vector, amplitude and frequency modulation, in time. C. Two linear kernels are fitted to the responses of the unit to the first 200 vocalizations - one for the amplitude (blue), and one
for frequency modulation (purple). D. A single joint non-linear response function is fitted as an exponential, to the firing rate of response versus the sum of the two linear predictions from the amplitude and frequency modulation vectors. E. Under STRF LNM, the stimulus is represented by its spectrogram. F. STRF. G. Non-linear function is fitted as an instantaneous transfer function to the firing rate of responses versus the linear prediction based on the STRF. H. Prediction for firing rate of responses to the remaining 150 vocalizations. Red – GLM model prediction, Green – STRF LN model prediction, Black – measured firing rate.

Over the population of neurons (all units N = 397, single units N = 172), the GLM model predicted responses to a novel stimulus sequence accurately (Figure 9A, prediction accuracy up to 0.8) and significantly higher than the spectrogram-based model (Figure 9B, C, paired t-test, p < 1e-18 all units, p < 1e-8 single units). Across all recorded units, the mean coefficient of correlation between the GLM model and recorded firing rate was 0.22 (standard error of the mean 0.01). This value was not significantly different from that for single units alone (0.19, standard error of the mean 0.01). 28% of all neurons (21% of single units) fitted to the GLM attained prediction accuracy higher than 0.3. The mean prediction accuracy of the spectrogram-based model was 0.17 for all units, and 0.15 for single units (standard error of the mean, 0.01). Only 17% of all units (14% of single units) exhibited prediction accuracy higher than 0.3. Thus, the responses to the vocalizations were accurately predicted by the reduced GLM, and this prediction was more accurate than the prediction given by a spectrogram-based model.
GLM prediction depends on the tuning of A1 units to ultra-sonic frequencies

We next compared the prediction accuracy of the GLM to the spectral and frequency modulation tuning properties of the units. There was a significant correlation ($r = 0.27$, $p < 1e^{-5}$, $N = 283$, Figure 10A) between the best frequency of the neuron and the model prediction accuracy across all units, but this correlation was not significant for single units alone ($p > 0.05$, $N = 125$). High prediction accuracy was exhibited by units whose BF was situated in the ultra-sonic range of the vocalization. 54% of all units (28% of single units) whose best frequency was above 40kHz exhibited prediction accuracy of 0.3 or higher. Their mean prediction accuracy was 0.33 (0.24 for single units). Nevertheless, 23% of units (21% of single units), for which BF was below 40kHz, exhibited a prediction accuracy of 0.3 or higher. Their mean prediction accuracy was 0.20 (0.21 for single units alone).
The model prediction accuracy was also significantly correlated with the overlap of the units' spectral receptive field and the power spectrum of the stimulus (Figure 10B, \( r = 0.37, p < 1e^{-10} \)) for all units, but not for single units (\( p > 0.05 \)). These findings demonstrate the best frequency plays a role in driving the responses of the units to the USVs, yet some units that are not tuned to the frequency range of vocalizations still exhibit predictable responses to USVs.

We also compared the model prediction accuracy to the FM sweep response properties of neurons. The model prediction accuracy was highly correlated with the FM sweep response strength (\( r = 0.66, p < 1e^{-7}, N = 74, r = 0.45, p < 0.01, N = 32 \) single units Figure 10C) and significantly correlated with the FM rate tuning index (\( r = 0.37, p < 0.005, \) all units, Figure 10D) for all units, but not for single units alone. The correlation between the response strength to USVs and the FM rate tuning index (Figure 7D) is consistent with the observed correlation between the FM rate tuning index and the model prediction accuracy (Figure 10D), as the FM temporal filters are typically restricted to a specific subset of FM rate transitions (Figure 8C). However, the FM directionality index did not correlate significantly with the model performance accuracy (\( p > 0.05, \) Figure 10E).
Figure 10. Correlation between the model prediction accuracy and neuronal tuning properties. A. GLM prediction accuracy versus best frequency for each unit. B. GLM prediction accuracy of each unit versus the normalized overlap between its spectral response profile and the power spectrum of USVs. C. GLM prediction accuracy versus response strength to FM sweeps. D. GLM prediction accuracy versus response strength to FM rate tuning index. E. GLM prediction accuracy versus FM directionality index.

A1 multi-units respond stronger to original as compared to transformed USVs

One of the hallmarks of vocalization coding is the differential sensitivity of neurons to original and transformed vocalizations (Wang et al. 1995). Reversing a complex sound preserves all the first-order statistical features of the sound, yet the higher-level features are modified...
(second-order features are temporally reversed). For example, an upward frequency sweep, when reversed, becomes a downward frequency sweep; however, the spectral amplitude distribution and contrast remain unchanged. As coding of the stimuli takes into account more complex features, reversing the vocalization is expected to evoke a different response than the original signal.

To test whether A1 responds more strongly to the original, rather than to transformed vocalizations, we presented a set of transformed vocalization sequences, in which the vocalizations were (1) accelerated (temporally compressed) by a factor of 1.5; (2) slowed down (temporally dilated) by a factor of 0.67; (3) reversed (see Methods, Figure 3). The responses of a representative unit to the four versions of the stimulus, transformed in time to compensate for the changes in the stimulus, are depicted in Figure 11. Note that some of the peaks in the responses of the unit persist across the transformations of the vocalizations (indicating features in the stimulus that drive the unit regardless of the transformations), whereas other firing rate peaks do not repeat across transformation of vocalizations. A similar pattern of responses would be expected based on the GLM of neuronal responses: Some features of USVs, when transformed, and convolved with a linear filter (Figure 8B) would be expected to result in a similar amplitude of activation, as those resulting from original USVs, whereas other USV features would not.
Figure 11. Time course of a representative unit to sequences of original and temporal transformed USVs. The firing rate of a representative unit in response to a segment of the USV sequence: black trace – responses to original USVs, cyan – compressed, green – dilated, red – reversed. To enable the comparison of which features of the stimulus the unit likely responded to, the time axes were transformed to compensate for the transformation imposed on the stimulus. For compressed, the time axis is expanded by x1.5, for dilated condition, the time axis is...
compressed by x0.67, for reversed, the time axis is reversed and singe the order of USVs was also reversed, the last segment is taken. These transformations, if applied to the sound waves, would have rendered hem similar to each other.

Across the population of cells, the response strength of units was lower for the transformed, as compared to the original USVs (N = 144, p < 0.05, Figure 12A), while the mean firing rate did not change (repeated measure ANOVA, p > 0.05, Figure 12B). The difference in the response strength was not due simply to changes in the baseline firing rate or standard deviation, which did not exhibit significant changes across stimulus types (repeated measure ANOVA, p > 0.05). For single units alone, the same trend was apparent (lower mean response strength for transformed, as compared to the original vocalizations), but the differences in response strength were not statistically significant (N = 50, p > 0.05 following Bonferroni correction). The lack of change of the firing rate was also predicted by the GLM fitted on the original vocalizations (repeated measure ANOVA, p > 0.05). These results demonstrate that the original vocalizations elicited stronger response strength, although not a greater firing rate, as compared to the transformed vocalizations, in A1 neurons.
Figure 12. A1 exhibits preference for original over temporally transformed vocalizations. A. Mean response strength of units to original, compressed (accelerated), dilated (slowed down) and reversed vocalizations. B. Mean firing rate in response to original, compressed (accelerated), dilated (slowed down) and reversed vocalizations. C. Mean prediction accuracy for responses to original, compressed (accelerated), dilated (slowed down) and reversed vocalizations. Solid – fitted on responses to original vocalizations; hatched – fitted on responses to transformed vocalizations. A-C: Error bars: standard error of the mean. Stars denote significance level in a repeated measures ANOVA after Bonferroni multiple comparisons correction relative to original (***: p < 0.001, *: p < 0.05). D. Prediction accuracy for all units for responses to compressed versus original USVs. Each dot denotes a unit. E. Prediction accuracy for all units for responses to dilated versus original USVs.
GLM predicts A1 neuronal responses less accurately to transformed than to original vocalizations

We next examined the accuracy of the fit of the GLM for responses to transformed USVs when fit on the original USVs. We first fitted the predictions of the GLM on the first 200 original USVs and used these fits to generate a prediction for USV responses to the remaining 150 transformed vocalizations. The model predictions were less accurate for the reversed, compressed and dilated vocalizations (N = 144, all units, p < 0.05, repeated measure ANOVA, Bonferroni multiple comparison corrected, same trend, but comparison nor significant for single units alone, Figure 12C).

However, the reduced model prediction accuracy change could potentially be due simply to under-sampling of the stimulus space during the model fit. To verify that it is indeed the prediction accuracy is indeed decreased due to a change in the response parameters, we fitted the model on the first 200 vocalizations of each of the transformed stimuli and tested it on the remaining 150 under each condition. After this change in analysis, the decrease in prediction accuracy was still significant for the dilated and compressed vocalizations (Figure 12C, D, E, p < 0.05, following Bonferroni multiple comparison correction, all units combined, but not significant for single units alone). The differences of processing of stimuli following temporal dilation and compression point to a differential processing mechanism for original and temporally transformed vocalizations. However, for reversed vocalizations, the prediction accuracy was consistent with that of the original stimuli (difference was not significant at p < 0.05 following Bonferroni correction), when the GLM was fitted on the reverse vocalizations. This is consistent with the explanation that the original stimuli under-sampled the statistical
space of reverse stimuli. Combined, these results indicate that the integration of the amplitude and dominant frequency modulation of the USVs best predicts the responses of A1 neurons to original vocalizations, but that the computations underlying this integration favor the natural statistics of the vocalizations.

Discussion

Rats communicate with ultra-sonic vocalizations (USVs). Yet despite their behavioral prevalence, little is known about how ultra-sonic vocalizations are encoded in the auditory system. Here, we characterized the responses of neurons in the awake rat primary auditory cortex (A1) to distinct USVs. A1 neurons exhibited significant responses to a subset of USVs. We found that the response strength of neurons in A1 to USVs was correlated with their best frequency, FM sweep response strength and FM rate tuning index. We constructed a reduced parameter predictive model, which relied on temporal integration of dominant frequency modulation and amplitude of the signal. This model accurately predicted the responses of A1 neurons to previously unheard vocalizations. These results contribute a significant advance over previous work on encoding of con-specific vocalizations in A1 by demonstrating a simple, yet precise computation that underlies their encoding. We also present evidence for preference in A1 responses to the temporal statistics of the original USVs: A1 responses had higher response strength, and the prediction of the model was more accurate for original, as compared to temporally transformed vocalizations. Our findings suggest that the neuronal circuits of processing ultra-sonic sounds are tuned to the ethologically relevant stimulus statistics.
Preference of A1 responses for temporal structure of original USVs

The hierarchical theory of cortical processing (Felleman and Van Essen 1991; Mineault et al. 2012; Zeki and Shipp 1988) posits that neural networks in more central auditory cortical areas encode progressively more complex features of the stimulus, increasing their preference for complex, specific acoustic objects, such as vocalizations (Hackett 2011; Kaas and Hackett 1998; Leaver and Rauschecker 2010; Recanzone 2008; Rouiller et al. 1991; Ter-Mikaelian et al. 2007). Indeed, for some species, including the non-human primate, it has been shown that neurons in A1 respond more strongly to the original rather than temporally transformed (compressed or dilated) vocalizations (Wang 2000).

The auditory system also exhibits tolerance to temporal transformations of acoustic stimuli. In humans, speech comprehension does not degrade perceptual accuracy for as much as a two fold compression (Beasley et al. 1980; Foulke and Sticht 1969). At the neurophysiological level, it has been shown that cells in A1 exhibit envelope following of the stimulus (Ahissar et al. 2001; Gehr et al. 2000; Gourevitch and Eggermont 2007; Mukamel et al. 2011; Nourski et al. 2009), thereby preserving their firing rate to original and temporally transformed stimuli.

Our results argue for a preference of neuronal responses in A1 to original over temporally compressed or dilated vocalizations. In evaluating the response strength of A1 neurons to different vocalizations, we found that the response strength, characterized by the relative variation in the firing rate as compared to the baseline was greater for original than for temporally dilated or compressed USVs (Figure 12A). On the other hand, the absolute mean firing rate during stimulus presentation did not vary significantly across the conditions (Figure
12B), which may be due to a mechanism that serves to maintain a constant response rate in neurons under different statistics of the stimulus, such as divisive normalization.

Our predictive model fits revealed a further statistically significant dependence of responses on the temporal structure of the stimulus. When a long vocalization sequence was transformed in temporal statistics, the accuracy of the predictive model of neuronal responses decreased significantly (Figure 12C). Thus, the response properties of neurons depended on the statistical makeup of the stimulus features. This dependence was not likely to be due to a simple temporal dilation or compression of the temporal kernel, as refitting the model on the transformed vocalizations preserved the decrease in the accuracy of the model's performance. The accurate fit of the predictive model for responses to original USVs demonstrate that the responses in A1 neurons reflect a computation, based on integration of the time course of frequency modulation and the amplitude of the dominant spectral component. The decrease in the accuracy of the prediction of the model for responses to transformed vocalizations suggests that the parameters of this computation are tuned to the statistics of the original, rather than transformed, USVs. Our model identifies a simple mechanism that likely leads to A1 responses, and allows for comparison of responses across stimulus types with varying statistical structure. Future studies are needed to determine whether the preference for original USVs by A1 neurons results in increased coding efficiency within this ethologically relevant statistical regime, as predicted by the efficient coding hypothesis (Barlow 1961).
**Similar computation underlying responses to reversed vocalizations**

Previously, it has been found that A1 neurons exhibit a stronger response to the original vocalizations than to their reversed versions in non-human primates (Averbeck and Romanski 2006; Romanski and Averbeck 2009; Wang and Kadia 2001; Wang et al. 1995). This preference for original vocalizations, which may partially be explained by differences in spectro-temporal structure of the original and reversed vocalizations, has been taken as a hallmark of vocalization specificity. However, other studies did not find a similar preference for the original over reversed vocalizations (Huetz et al. 2009; Schnupp et al. 2006). To test whether the high accuracy in response prediction of the neurons was specific to the original vocalizations, we presented a stimulus sequence consisting of reversed vocalizations (Figure 3D). Like with temporally transformed vocalizations, we found that response strength of the neurons, but not the mean firing rate, was decreased during the presentation of reversed USVs as compared to the original (Figure 12A, B). Unlike for temporally transformed USVs, the reduced response strength was not accompanied by reduced model prediction accuracy, as would be predicted were the neurons to respond to the reversed USVs via a differential mechanism (Figure 12C). These results suggest that the original and reversed vocalization stimuli, which overlap in their temporal and frequency modulation spectrum, covering the same range (Figure 3), appear to activate A1 via a similar computational mechanism.

**Response strength of A1 neurons to USVs**

A previous study in the mouse A1 found that neurons, located in the sonic range of the tonotopic axis, were responsive to ultra-sonic tones, even if their best frequency was found in the sonic range (Linden et al. 2003). We also found that many neurons whose best frequency
was outside the ultra-sonic range responded significantly to USVs. However, the responses
neurons whose best frequency overlapped with the power spectrum of the USVs were more
accurately predicted by the GLM – suggesting that their responses are likely controlled by a
reduced set of precise computations, as evidenced by our predictive model.

Our results extend the previous characterization of responses to a different type of
ultra-sonic vocalizations in the mouse (Holmstrom et al. 2010; Lin and Liu 2010; Liu and
Schreiner 2007; Liu et al. 2006). A subset of A1 neurons in the mouse exhibit precise temporal
following and short response onset latency in response to ultra-sonic infant pips (Lin and Liu
2010). Our GLM may provide for the computation that produces the reduced onset latency:
reduced response onset latency may be predicted by our GLM, if both frequency modulation
and amplitude kernels were to exhibit a short time to peak. The correlation of the identity of the
neurons with their response strength and predictability to USVs remains to be experimentally
tested.

*Neuronal correlates of GLM performance*

The GLM was partially based on integration of frequency modulation and amplitude of
the dominant spectral channel over time. Not surprisingly, the response strength to the USVs in
A1 was correlated strongly with response strength to FM sweeps. Therefore, the mechanisms
that underlie the computation identified in the GLM may share the neuronal circuits with those
giving rise to FM responses. FM responses in A1 have been proposed to have both subcortical as
well as cortical origins (Atencio et al. 2007; Nelken and Versnel 2000; Phillips et al. 1985;
Shamma et al. 1993; Trujillo et al. 2011; Zhang et al. 2003). A recent study demonstrated that
direction tuning of A1 responses may be explained by integration of the incoming signals from the ascending afferent auditory pathway, that are already directionally tuned (Kuo and Wu 2012). These signals may originate as early as the inferior colliculus, a brain area peripheral to A1, where neurons are tuned to specific direction of frequency modulation of the incoming sound (Clopton and Winfield 1974; Hage and Ehret 2003; Rees and Moller 1983; Suga 1968).

However, additional mechanisms that generate frequency modulation tuning, including sideband facilitation and inhibition (Razak and Fuzessery 2008; 2006), may also be at play in driving the responses to USVs within A1. Sideband facilitation and inhibition may also be behind the responses to integration of amplitude modulations of those neurons that do not exhibit tuning to ultra-sonic tone pips. These mechanisms, carried out by intra-cortical connections, have previously been shown to give rise or modulate responses in A1 neurons to signals outside the center of their frequency response area (summarized in Schreiner et al. 2011; Sutter 2005), and may drive responses of A1 neurons to amplitude modulations in USVs.

The low-parameter description of the signal allowed us to represent the sound with 1 ms precision, and preserve high frequency resolution in representing the frequency modulation, while implementing the computationally intensive maximum likelihood optimization technique. To arrive at this representation, the signal was transformed through a non-linear operation (preceding the fit of the GLM): extraction of the maximum instantaneous frequency of the signal. It is plausible that the rat auditory system implements a similar operation in isolating the vocalization from the background noise. The neuronal correlate of this operation would involve a potential winner-take-all circuit, implemented through, for example, lateral inhibition across direction-tuned neurons (Xie et al. 2002).
These mechanisms may generalize to more complex vocalizations which contain multiple spectral components such as harmonics, and to environmental sounds, which contain broadband components (Geffen et al. 2011). Our model may be extended to include an intra-cortical lateral inhibitory circuit that would detect the maximally activated direction-tuned channels. This mechanism would allow for encoding of multiple dominant spectral components, separated across several channels. Signal in each channel would then undergo a processing cascade, described by a GLM, based on frequency modulation and amplitude of each distinct channel. Further studies of processing of harmonic vocalizations are needed to determine whether the computation proposed in this paper generalizes to harmonic and environmental sounds.

In this work, we explored the computation that underlies the responses of neurons in A1 to USVs. We found that the computation relies on integration of the time course of frequency modulation and amplitude of the dominant spectral component of the USVs. The use of USVs as stimuli enabled us to develop a simple, yet powerful model of A1 responses. A similar reduction of the parameters of the stimulus may prove powerful in predicting and understanding the mechanisms of A1 responses for other types of stimuli, including speech and music. The predictability and precision of responses of individual neurons further corroborates the role of A1 as a brain area important for extracting behaviorally meaningful acoustic information.
2: EMERGENCE OF INVARIANT REPRESENTATIONS OF VOCALIZATIONS

Overview

An essential task of the auditory system is to tell apart different communication signals, such as vocalizations. In everyday acoustic environments, the auditory system needs to capable of discriminating between vocalizations under different acoustic distortions. To achieve this, the auditory system is thought to build a representation of vocalizations that is invariant to 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 non-primary auditory cortex of rats to original and acoustically distorted vocalizations. We found that populations of neurons in the non-primary 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.

Introduction

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 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 2012;
2010). Here, we tested whether populations of neurons beyond A1, in a non-primary 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
supra-rhinal auditory field (SRAF) (Arnault and Roger 1990; Polley et al. 2007; Profant et al.
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 non-primary 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.

Rat communication signals (ultra-sonic vocalizations, or 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. While we do not yet have an understanding of which USVs have a distinct behavioral meaning for rats, the present study identified 8 USVs with distinct spectro-temporal properties, and constructed their acoustic distortions along the dimensions that are essential for their encoding in the auditory pathway. We recently found that the 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 amplitude-modulation of the dominant spectral component (Carruthers et al. 2013). Therefore, we used these statistical 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 identify, 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.
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 weeks of age. Rats were housed in a temperature and humidity-controlled vivarium on a reversed 24 hour light-dark cycle with food and water provided ad libitum.

Stimuli.

The original vocalizations were extracted from a recording made when two adult male rats, housed in isolation, were placed in a single cage together for 2 hours. Vocalizations were recorded using a free-field ultra-sonic microphone (Avisoft Bioacoustics, CM15, sensitivity 50 mV/Pa, frequency range: 10 - 200 kHz, input-referred self-noise level 18 dB). 8 representative USVs were drawn from the recorded set. These USVs were then parameterized and purified following methods published previously by our group (Carruthers et al. 2013). Briefly, a noiseless version of the vocalizations was constructed, using an automated procedure, by isolating the dominant frequency and amplitude for each noisy vocalization. The noiseless signal was constructed 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, using the relation

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

For each of these 8 original vocalizations we generated 8 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 played at a rate of 2.5Hz.

**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), temporal dilation (T+, length x 1.25), spectral compression (FM-, bandwidth x 0.75), spectral dilation (FM+, bandwidth x 1.25), spectro-temporal compression (T-/FM-, length and bandwidth x 0.75), spectro-temporal dilation (T+/FM+, length and bandwidth x 1.25), center-frequency increase (CF+, frequency + 7.9 kHz), and center-frequency decrease (CF-, frequency – 7.9 kHz). Spectrograms of original vocalizations and transformations of one of the vocalizations are shown in Figure 10.

**Microdrive implantation.**

Rats were anesthetized with an intra-peritoneal injection of a mixture of ketamine (60 mg per kg of body weight) and dexmedetomidine (0.25 mg per kg). Buprenorphine (0.1 mg/kg) was administered as an operative analgesic with Ketoprofen (5 mg/kg) as post-operative analgesic. A small craniotomy was performed over A1 or SRAF. 8 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 40mm 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 (10mAmp, 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; 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 inch 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 3dB. Spectral and temporal distortion products as well as environmental
reverberation products were >50 dB below the mean 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 40mm between sessions to avoid repeated recording from the same units. Tetrode position was noted to ±20mm precision. Electo-physiological data from 24 channels were filtered between 600 and 6000 Hz (to obtain spike responses), digitized at 32kHz 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) it averaged at least 0.1 Hz firing rate during stimulus presentation, and 2) the unit’s spike-count during the presentation of a single vocalization conveyed at least 0.058 bits of information about the vocalization identity. This estimation was computed 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. 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:

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

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 milliseconds. 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 of 100 x N matrix of responses for each of the 72 vocalization/transformations (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 Generalization 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 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 performance 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.

Generalization.

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 generalization, we trained and tested on vectors drawn from presentations of one transformation and from the original vocalizations. When testing all-transformation generalization, we trained and tested on vectors drawn from all 9 transformation conditions.

Within-transformation Performance.

For each subset of transformations on which we tested generalization performance, we also trained and tested classifiers on each individual transformation condition. We refer to performance of these classifiers, averaged over the transformation conditions in the corresponding generalization task, 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 define the generalization penalty as the difference between the within-transformation performance and the generalization 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 detailed in depth in Chapter 1.

We selected 8 distinct vocalizations from recordings of social interactions between male rats. We then generated 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 Figure 10a, and Figure 10b 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 (Figure 11). SRAF units exhibited similar patterns of responses (Figure 12). For instance, the representative A1 unit shown in figure 11 responded significantly to all of the original vocalizations except vocalizations 5, 6, and 8 (row 1). Meanwhile, the representative SRAF unit in figure 12 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.
Figure 13. 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. Original: original; T-: temporally compressed by factor of 0.75; T+: temporally stretched by factor of 1.25; CF+: center frequency shifted up to 7.9 kHz; CF-: center frequency shifted down by 7.9 kHz; FM+: frequency modulation scaled by a factor of 1.25; FM-: frequency modulation scaled by a factor of 0.75; T-/FM-: temporally compressed and frequency modulation scaled by a factor of 0.75; T+/FM+: temporally stretched and frequency modulation scaled by a factor of 1.25.
Figure 14. Peri-stimulus-time histograms of an exemplar A1 unit showing selective responses to vocalization stimuli. Each column corresponds to one original vocalization, and each row to one transformation of that vocalization. Histograms were first computed for 1ms time-bins, and then smoothed with 11-ms hanning window. The grey highlight shows the duration of the vocalizations.
Figure 15. Peri-stimulus-time histograms of an exemplar SRAF unit showing selective responses to vocalization stimuli. Each column corresponds to one original vocalization, and each row to one transformation of that vocalization. Histograms were first computed for 1ms time-bins, and then smoothed with 11-ms hanning window. The grey highlight shows the duration of the vocalizations.
To compare the responses of populations of units in A1 and SRAF, we selected subpopulations of units that were matched for firing rate distribution (Figure 13A). 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) (Figure 13B). 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 (Figure 13C, 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 (Figure 13D), but this difference was not significant (Kolmogorov-Smirnov test, p = 0.084).
Ensembles of A1 and 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. 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 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.

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 (Figure 13E, 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 13G, H shows the classification performance broken down by vocalization for pairwise classification for A1 (Figure 13G) and SRAF (Figure 13H). There is high variability in performance between vocalization pairs for either brain area. However, the performance levels are 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 acoustically. Therefore, we trained and tested classifiers for vocalizations based on population neuronal responses and compared their performance under within-transformation and generalization conditions (Figure 14A). In within-transformation condition, the classifiers were trained and tested to discriminate responses to vocalizations under a single transformation. In generalization condition, the classifier was trained and tested in discriminating responses to vocalizations in original form and one or all transformations. The difference between within-transformation and the generalization classifier performance was termed the generalization penalty. If the neuronal population exhibited low invariance, we expected the generalization performance to be lower than within-transformation performance and the generalization penalty to be high (Figure 14A top). If neuronal population exhibited high invariance, we expected the generalization performance to be equal to within-transformation performance and the generalization penalty to be low (Figure 14A bottom).
Within-transformation
A1: High discriminability

Generalization
A1: Low discriminability

SRAF: High discriminability

Within-transformation
A1: High discriminability

Generalization
A1: Low discriminability

SRAF: High discriminability

B Pairwise, Per-Transformation

C Pairwise, All-Transformation

D 8-Way, Per-Transformation

E 8-Way, All-Transformation
Figure 17. Classifier performance on generalization and within-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 vocalizations for original and transformed vocalizations simultaneously. Predictions of the hierarchical coding model: Generalization classification performance is poor for A1, demonstrating less invariance, and high for SRAF, demonstrating higher invariance. B, C) Performance when discriminating each vocalization from one other vocalization (pairwise classification). D, E) Performance when discriminating each vocalization from all others (8-way classification). B, D) Performance when generalization is performed across the original vocalizations and one transformation at a time (per-transformation generalization). C, E) Performance when generalization is performed across all eight transformations and the originals at once (all-transformation generalization).

To ensure that responses to a select transformation were not skewing the results, we computed generalization both for each of the transformations and for all transformations. In per-transformation generalization condition, the classifier was trained and tested in discriminating responses to vocalizations in original form and under one other transformation. In all-transformation generalization 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 generalization condition as compared to within-transformation condition than neuronal population in SRAF. Figures 14B,C and 6 present the comparison between generalization 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 generalization condition (Figure 14, B-E). However, the difference in performance between within-transformation and generalization conditions was higher in A1 than in SRAF: SRAF populations suffered a smaller generalization penalty under all conditions tested (Figure 15). This effect was present under both pairwise (Figures 14B, C, 15A, B) and 8-way classification (Figures 14D, E, 15C, D), and for generalization in per-transformation (Figures 14B, D, 15A, C, pairwise classification, p = 0.028; 8-way classification, p = 1.9e-4; Wilcoxon paired sign-rank test) and all-transformation mode (Figures 14C, E, 15B, D; pairwise classification, p = 1.4e-5; 8-way classification, p = 0.025; Wilcoxon paired sign-rank test). 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.
Figure 18. Generalization penalty (difference between within-transformation performance and generalization 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, B) Generalization penalty when discriminating each vocalization from one other vocalization (pairwise classification). C, D) Generalization penalty when discriminating each vocalization from all others (8-way classification). A, C) Generalization penalty when generalization is performed only across the original vocalizations and one vocalization at a time (per-transformation generalization). B, D) Generalization penalty when generalization is performed across all eight transformations and the originals at once (all-transformation generalization). * p<0.05; *** p<0.001.
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 non-primary 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 (Figure 10). We found that, while neuronal populations in A1 and SRAF exhibited similar selectivity to vocalizations (Figures 11, 12, 13), neuronal populations in SRAF exhibited higher invariance to acoustic transformations of vocalizations than in A1 (Figure 14, 15). 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 ultra-sonic vocalizations can be parametrized 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 timecourse 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 ultra-sonic 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.

A1 neurons adapt to the statistical structure of the acoustic stimulus (Asari and Zador 2009;
Blake and Merzenich 2002; Kvale and Schreiner 2004; Rabinowitz et al. 2013; Rabinowitz et al.
2011). We previously found that in encoding ultra-sonic vocalizations, A1 neurons not only
responded more strongly to original rather than to temporally transformed USVs, but their
responses could be predicted by a simple model with high accuracy (Carruthers et al. 2013).
Therefore, the amplitude of frequency shift and frequency modulation scaling coefficient were
chosen on the basis of the range of the statistics of ultra-sonic 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 types of transformations that were applied to the stimulus, and these results 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. 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.

There exist multiple mechanisms that could potentially explain the increase in invariance we observe between A1 and SRAF. As previously suggested, simple cortical microcircuits in A1 can transform incoming responses into a more feature-invariant form (Atencio et al. 2009). Such a transformation could provide a mechanism for decreasing the sensitivity of neural responses to feature-based stimulus transformations. Alternatively, higher auditory brain areas may be better able to adapt to statistical structures in auditory stimuli (Rabinowitz et al. 2013). Such 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 auditory cortex, the observed changes may already be encoded within different groups of neurons or within different cortical layers within the primary auditory cortex. Further investigation includes 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.
3: BEHAVIORAL MEASUREMENTS OF VOCALIZATION DETECTION

Overview

While we can learn a great deal about the information-processing ability of the brain by looking directly at neural responses, the ultimate purpose of any neural processing system is to influence behavior. No matter how much information we believe is present in our neural recordings, that information only matters to the organism if it can influence behavioral action. For example, recent studies have shown that information that appears to be present in the neural code might not be utilized in behavior (Carney et al. 2014). In particular, if we wish to know whether a rat can perceive the presence of an auditory stimulus such as a rat vocalization, it is not sufficient to show that the rat’s neural responses are significantly different when the stimulus is present versus when it is absent; we must show that the rat is actually capable of taking different actions based on whether the stimulus is present.

With this in mind, we set out to develop an experimental procedure to probe rats’ ability to perceive differences in ultrasonic stimuli, and in USVs in particular. The aim of this project was to design, build, and employ an experimental setup which could be used to train rats to make decisions based on auditory stimuli, track those decisions, and dynamically adjust the stimuli according to a predefined protocol. We used this experimental setup to measure psychometric properties of rat audition, and in particular to measure the ability of rats to detect the vocalizations of other rats with random noise superimposed.

In order to get the rats to report their experiences we had to both teach them how to make reports, and motivate them to do so. Past behavioral work on rodents has used either food or
water to motivate the rodent’s engagement in a task. We elected to use water as a reward due to the simplicity of the delivery mechanism, and for the fact that it could be delivered in arbitrarily small doses, allowing us to keep the rats motivated for longer periods.

Briefly, we deprived our rats of water for approximately 24 hours prior to conducting our experiments. We then placed a rat in a cage containing three nose-ports (Figure 16). Our software system could detect when the rat inserted his nose into each nose-port, and would trigger the delivery of water when the rat poked the right port at the right time. By slowly raising the complexity of the motions the rat had to go through in order to receive the reward, we taught the rats to hold their nose in one port until they heard a target stimulus (an ultrasonic trill in the same range as natural vocalizations), and only then to go to another port to claim a reward. With the rats so trained, we could superimpose noise over the target stimulus and see how different levels of noise affected the rate at which the rat would claim rewards.

System Design

We had to take into account a number of physical and practical constraints when designing the experimental test cage for this project. Naturally, in order to accurately test audition, the experiment had to be housed in a noise-controlled environment. Additionally, as rats are nocturnal animals, we expect them to remain more alert, and to perform better on tasks when in darkness. To satisfy both of these constraints, the animal cage and supporting equipment were housed in a dark, double-walled sound booth.

While rats perform best in darkness, we considered it necessary for many practical reasons that we be able to visually observe the rats while they performed the tasks. Fortunately, rat vision
does not extend as far into the low-frequency end of the spectrum as does human sight. Equally fortunately, many commercial web-cameras are capable of picking up infra-red light. This allowed us to place an infra-red lamp inside the sound booth, confident that the rat would not find the dull red glow disturbing, and watch the rat via camera.

The cage was set up, as shown in figure 16, so that the only objects within reach of the rat were the nose-ports through which the rat was to interact with the experiment. This decision was made both to protect the lab equipment (as rats can by quite destructive when bored), and to encourage the rat to investigate the nose-ports, and thus discover the rules of the experiment. The speaker and camera were positioned above the cage, and the water reservoir was set on top of the cage, again making sure that the wires and water tubes did not come within reach of the rat. The camera was angled so as to clearly show the nose-ports, so that the experimenter could recognize if the rat’s interactions with the ports were inconsistent with the behavior being recorded.

The nose-ports themselves were designed to allow the rat to easily and intuitively interact with the experiment. The ports were machined from cylinders of Delrin® plastic, which is strong enough that it is difficult for the rats to damage. A conical recess was carved into each cylinder to form a space where a curious rat would be tempted to poke its nose. A hole was drilled in the back of the recess for the water-tube, positioned below the center of the port so that a pointy-nosed animal could drink from it conveniently. Finally, four evenly spaced holes were then drilled around the edges of the recess, perpendicular to the edge of the cylinder, forming two opposing pairs. For each pair of holes, one hole housed an infra-red LED, and the other a photo-
transistor sensitive to infra-red light. The result, shown in figure 17, was a crossed pair of infra-red beams, each falling incident on a different photo-transistor (Figure 17).

Figure 19. General layout of experimental setup. The cage is positioned within a double-walled sound-booth, and left in darkness.
Figure 20. Construction of the nose-ports. The nose ports are machined from a solid plastic cylinder, with a conical cavity as shown in the left panel. Two photo-transistors and two IR-LEDs are embedded in the cylinder as shown, such that the two beams cross the cavity to fall on the two transistors as shown in the right panel.

For each nose-port, the two photo-transistors were connected in series, and a 5V potential applied over them (Figure 18). In the resulting circuit, a lack of light falling on either photo-transistor would result in a broken circuit. This allowed us to detect an intrusion into the nose-port as long as either beam was broken, which was sufficient to accurately detect the presence of a rat’s nose.
The water delivery system was fed by a gravity-assisted reservoir, and controlled via high-speed solenoid pinch-valves. With the reservoir set above the cage, water would naturally flow down the tubing and out through the holes set in the nose-ports. To control this flow, we set a separate pinch-valve on the tube leading to each nose-port. The valves remained closed by default, meaning that the rat did not receive water in the event that the power-source became disconnected or accidentally turned off.

The electrical components of the system were connected to and controlled via a high-speed digital-to-analog card (National Instruments). Each nose-port required two digital channels to operate: one to detect whether the beams in that port were blocked, and one to control water delivery through that port. The photo-transistors inside the nose-ports were connected via a simple circuit (shown in figure 18) to a digital input line on the N.I. card. Similarly, the pinch-valve controlling the flow of water to the port is connected to a digital output line via the circuit shown in figure 19. The valves required a higher voltage and more power than could be supplied
by the N.I. card, and so they were connected to a separate 12V power source also housed within the sound-booth. The additional complexity of the pinch-valve circuit was necessary to isolate the N.I. card from potential inductive voltage spikes incurred as the valves were turned off.

Figure 22. Circuit used to control the solenoid pinch-valves. As the valves are actuated by solenoids, we expect that they will have significant inductance. To prevent any inductive voltage spikes from damaging the data acquisition card, we use an optocoupler to isolate the valve from any electrical contact with the digital equipment.

Experimental Protocol

Each rat was set onto a water-deprivation schedule at least two days before training began. Rats were given water *ad-libitum* for 45 to 90 minutes each day, at approximately the same time, and
denied water the rest of the day. Each rat’s weight was recorded daily for four days before water-deprivation began to establish a baseline weight. While on water-deprivation, each rat was weight daily, before being given water, to ensure that their weight did not fall below 80% of their baseline weight.

Once water-deprived, each rat began a program of association training. Each rat was placed in the cage alone. One nose-port was designated the challenge port, and another was designated the reward port. In the association training task, any ingress into the challenge nose-port was quickly rewarded with a dose of water in the reward port, and paired with a presentation of the reward stimulus. This program continued until the rat had learned to successfully receive several dozen rewards in quick succession. Rats typically achieved this level of familiarity with the equipment within the first two or three sessions, where each session lasted between 45 minutes and an hour.

Once the rats were performing adequately on the association training task, they were moved to a simple detection task. Within this task, rats had to hold their nose in the challenge port for some variable amount of time. Once they reached the target hold-time, the reward stimulus was played, signaling that the rat should go to the reward port. In this task, the rats only received a reward if they moved to the reward-port within a short time of the reward stimulus being played.

It bears mentioning here that the rats seemed to try their hardest to avoid learning the task. For instance, in this simple detection task, the rats would attempt to exploit the variable length of the required hold-time. They would learn the approximate length of the minimum required
hold-times, and would repeatedly hold the challenge port for that length of time and then attempt to retrieve their rewards. It seems that the rats found it easier to move back and forth between ports at fixed time-intervals than to listen carefully for the reward stimulus. In order to discourage this exploitative behavior, we implemented a number of additional procedures. If the rat attempted to claim a reward early, then we repeated the length of time that the rat was required to hold on the next trial. We also implemented a time-out in between trials, limiting the speed at which the rat could go through trials.

These mitigating procedures met with variable success, and parameters such as the length of the time-out, and the distribution of required hold-times had to be tuned frequently. The arousal state of the animal also seemed to be a confounding variable in levels of performance on the task: if the rat was upset from being handled and put into the cage, or if the rat had missed a day of training, then performance suffered significantly...

Once the rats were performing at acceptable levels on the simple detection task, they were moved on to a more complex version of the task. In this task, at the end of each required holding period, the rats were presented with either a frozen white-noise stimulus, the target stimulus, or the target stimulus with the same frozen noise superimposed (Figure 20). On trials where the rats were presented with noise only, they could receive a reward only if they continued to hold until the reward stimulus was played a fixed time later. On trials where the target stimulus was played with noise superimposed, or where the target stimulus was played first without any noise, they could claim the reward only if they reacted quickly. By playing the
target stimulus at different volumes we could probe the levels at which the rats were capable of detecting the vocalizations in the noise.

![Figure 23. Schematic timeline of two example trials of our discrimination protocol. In the first trial shows an example where the target stimulus is played first with noise superimposed, the second trial shows an example where a distractor stimulus is played first, followed by the target stimulus. Then rat must hold the challenge port during the period marked in red, and then has the opportunity to claim a reward during the period marked in blue.](image)

The software to manage this experiment had to meet a number of requirements. It had to be modified frequently as the experimental protocol was adjusted and updated. It had to be robust to the many ways that the rat might interact with the nose-ports, as the rats learned and tested the parameters of the system. It also had to present the rat with a consistent interface; any irregularities in how the experiment responded to the rat’s actions could potentially cause significant setbacks in the rat’s training.

In order to satisfy these requirements, we elected to develop the software system as a finite-state machine (FSM). A finite-state machine, as the name suggests, is a simple mathematical codification of a system in which the system is represented by a number of states, and the
allowed transitions between those states. In our system, each state was represented by a set of five actions: one to perform when each nose-port was intruded upon, one to perform when each nose-port was released, and one to perform on every iteration of the main loop. The transitions between states invoked external actions such as playing stimuli, turning water on and off, or writing to log-files.

For illustration, we can see a representation of our final experimental protocol in figure 21. When in state 1, the initial state before a trial has started, the only action that will trigger a transition is if the rat intrudes on the challenge port, which triggers a transition to state 2. From state 2, a transition is triggered if either the rat withdraws from the challenge port (which takes us to a time-out state before re-starting the trial), or if the main-loop action determines that we have been in state 2 long enough to play a stimulus.

By constraining our software system in this way, we greatly simplified the process of developing a new experimental protocol, or of modifying an existing one. Compartmentalizing our code, and minimizing the amount of information that had to be passed between states, we significantly reduced the number of changes that had to be made to the code each time the protocol was updated.
Figure 24. Schematic of the software model used for the discrimination task. Circles denote states in the finite state machine, and arrows denote allowed transitions. Each transition can trigger actions such as delivering water, or writing to a file.
Figure 25. Example session of the simple detection task, as the rat is just starting to learn the task. The horizontal axis denotes the amount of time that the rat would have had to hold the challenge port to get a reward, the vertical axis denotes how long he actually held for. Blue marks denote trials on which the rat received a reward. Red marks denote trials on which the rat either released too early, or simply failed to claim the reward.

Once the rats had learned the task, their accuracy and reaction times improved quickly. In figure 22 we can see an example of an early session with one particular rat, showing the time the rat would have needed to hold in order to receive a reward, compared to the time that the rat actually held the nose-port. We see that the length of time the rat is holding is essentially
independent of the length of time it should have held. In figure 23, we see a later session from the same rat. We see that by the time of the latter session, the rat has learned to wait until it hears the reward stimulus on the majority of trials, and reacts very quickly when it does hear the reward stimulus.

Figure 26. Example session of the simple detection task, by the same rat as in figure 22, but after being trained on the task for several weeks. The horizontal axis denotes the amount of time that the rat would have had to hold the challenge port to get a reward, the vertical axis denotes how long he actually held for. Blue marks denote trials on which the rat received a reward. Red marks denote trials on which the rat either released too early, or simply failed to claim the reward.
During each experimental session we recorded the number of trials that the rat initiated, the number of those trials where the rat successfully retrieved a reward, the number where the rat unsuccessfully tried to retrieve a reward, and the number where the rat did not try to retrieve a reward. By comparing the number of successful and unsuccessful attempts under the various conditions, we could measure some sense of the difficulty of the task under those conditions. For instance, we can see in figure 24 that all three rats (once trained) successfully retrieved a reward at rates approaching 100% when the target was played at 20dB above the noise background.

Figure 27. Psychometric curves for three different rats. Green and blue lines show the fraction of those trials where the rat attempted to receive a reward on which the rat was successful. Green lines show trials where the rat was presented with only the target stimulus first, blue lines show trials where the rat was presented with the target stimulus embedded in noise. The black horizontal line shows the fraction of trials where a distractor stimulus was played first in which the rat attempted to receive a reward after hearing the distractor.
**BIBLIOGRAPHY**


Pelleg-Toiba R, and Wollberg Z. Discrimination of communication calls in the squirrel monkey: "call detectors" or "cell ensembles"? *Journal of basic and clinical physiology and pharmacology* 2: 257-272, 1991.


Rust NC, and DiCarlo JJ. Balanced increases in selectivity and tolerance produce constant sparseness along the ventral visual stream. *J Neurosci* 32: 10170-10182, 2012.

Rust NC, and DiCarlo JJ. Selectivity and tolerance ("invariance") both increase as visual information propagates from cortical area V4 to IT. *J Neurosci* 30: 12978-12995, 2010.


