Relative salience of spectral and temporal features in auditory long-term memory

Pingbo Yin^{a)}

Neural Systems Laboratory, Institute for Systems Research, 2207 A.V. Williams Building,
 University of Maryland, College Park, Maryland 20742, USA

Shihab A. Shamma^{b)}

Neural Systems Laboratory, Institute for Systems Research, Electrical and Computer Engineering Department,
 2203 A.V. Williams Building, University of Maryland, College Park, Maryland 20742, USA

Jonathan B. Fritz

Neural Systems Laboratory, Institute for Systems Research, 2207 A.V. Williams Building,
 University of Maryland, College Park, Maryland 20742, USA

(Received 24 June 2016; revised 5 October 2016; accepted 9 November 2016; published online 2 December 2016)

In order to explore the representation of sound features in auditory long-term memory, two groups of ferrets were trained on Go vs Nogo, 3-zone classification tasks. The sound stimuli differed primarily along the spectral and temporal dimensions. In Group 1, two ferrets were trained to (i) classify tones based on their frequency (Tone-task), and subsequently learned to (ii) classify white noise based on its amplitude modulation rate (AM-task). In Group 2, two ferrets were trained to classify tones based on correlated combinations of their frequency and AM rate (AM-Tone task). Both groups of ferrets learned their tasks and were able to generalize performance along the trained spectral (tone frequency) or temporal (AM rate) dimensions. Insights into stimulus representations in memory were gained when the animals were tested with a diverse set of untrained probes that mixed features from the two dimensions. Animals exhibited a complex pattern of responses to the probes reflecting primarily the probes' spectral similarity with the training stimuli, and secondarily the temporal features of the stimuli. These diverse behavioral decisions could be well accounted for by a nearest-neighbor classifier model that relied on a multiscale spectrotemporal cortical representation of the training and probe sounds. © 2016 Acoustical Society of America.

[<http://dx.doi.org/10.1121/1.4968395>]

[MLD]

Pages: 4046–4060

I. INTRODUCTION

When animals learn, or are actively engaged in a behavioral task, the cortical representation of relevant sensory stimuli can adapt to optimize task performance (Polley *et al.*, 2007; Fritz *et al.*, 2003; David *et al.*, 2012; Yin *et al.*, 2014). In most studies of learning or attention-driven plasticity, target stimuli have been tones or other simple stimuli. However, most natural acoustic stimuli are complex sounds that have multiple feature dimensions such as tone frequency, amplitude and frequency modulation rates, and bandwidth. All of these features are represented and some are mapped in primary auditory cortex (A1) (Schreiner *et al.*, 2000). Nevertheless, the saliency of these features, alone or in combination, in auditory object recognition, remains uncertain (Yin *et al.*, 2010; Stilp *et al.*, 2010, 2012). For example, in an auditory short-term memory (STM) task in monkeys, it has been suggested that spectral features are more salient for discriminating complex acoustic stimuli (Fritz *et al.*, 2005; Scott *et al.*, 2012, 2013), and that

temporal similarity between auditory stimuli was not a significant factor in predicting error rate responses to a distractor during an auditory delayed match-to-sample task (Scott *et al.*, 2013). These results on auditory STM in monkeys are consistent with findings that persistence of auditory STM in starlings was significantly shorter for signals varying only in the temporal domain than for tonal stimuli varying only in the spectral domain (Zokoll *et al.*, 2008b). However, the relative salience of temporal features may become more important to an animal if frequency or pitch information is relatively impoverished in the stimulus set. In this study, we explore the relative salience of spectral and temporal cues in an auditory long-term memory (LTM) task with ferrets trained to recognize sounds defined in three zones along two feature dimensions: spectral (frequency) or temporal [amplitude modulation (AM)] or their combination. Other acoustic feature dimensions were held constant for all stimuli (sound location, duration, intensity, onset and offset envelopes were fixed). The study explored the following issues:

- (1) How do spectral and temporal acoustic features each contribute, singly and collectively, to the encoding and retrieval of stimuli stored in auditory LTM?

^{a)}Electronic mail: pyin@umd.edu

^{b)}Also at: Ecole Normale Supérieure, 29 Rue d'Ulm, Paris 75005, France.

- (2) Can the behavioral results from single feature auditory LTM tasks predict the relative perceptual weighting of spectral and temporal acoustic cues in LTM retrieval in more complex dual feature auditory stimuli?

II. METHODS

A. Subjects

Adult female ferrets (*Mustela putorius*) obtained from Marshall Farms (North Rose, NY) were used in these behavioral experiments. All ferrets in the study ($n = 4$) were over a year in age, and weighed between 600 and 900 g. They were housed in pairs in facilities accredited by the Association for Assessment and Accreditation of Laboratory Animal Care, and were maintained on a 12-h light–dark artificial light cycle. During behavioral training and testing, the animals were placed on a water-control protocol on which they had restricted access to water during the week except as a reward during behavior or as liquid supplements. The ferrets were trained 5-days per week and obtained *ad libitum* water over weekends. They were brought to the laboratory during daily training sessions and returned to the central animal facility after completion of behavioral testing. During behavioral training or testing days, animals obtained water primarily during task performance. Supplementary water was given after task performance if the animals did not obtain sufficient water during the behavioral session. The animal's health condition was monitored on a daily basis to avoid dehydration and weight loss (animals were maintained above 80% of *ad libitum* weight). All procedures were conducted in accordance with the National Institutes of Health's Guide for the Care and Use of Laboratory Animals (which was published by National Research Council of the National Academies), and were approved by the Institutional Animal Care and Use Committee of the University of Maryland.

B. Experimental apparatus

Ferrets were trained in a custom-built transparent Lucite testing box (18 cm width $\times$ 34 cm depth $\times$ 20 cm height), placed within a single-walled, sound attenuated chamber. A lick-sensitive waterspout (2.5 cm $\times$ 3.7 cm) stood 12.5 cm above the floor in front of the testing box, and animals could easily lick from the waterspout to obtain water through a small opening in the front wall of the box. The waterspout was connected to a computer-controlled water dispenser (Crist Instrument Co., Inc., Hagerstown, MD), and also to a custom-built interface box that converted the licks into a Transistor-Transistor Logic digital signal feed to a computer. A loudspeaker was positioned 20 cm in front of the testing box for sound delivery during behavioral training, and the animal's behavior was streamed with a video camera, displayed graphically trial-by-trial, and continuously monitored on a computer screen.

All stimuli used in the training were generated using MATLAB (The MathWorks, Natick, MA) at a 40 kHz sampling rate. The sounds were converted at 16-bit resolution through a NI-DAQ card, then amplified (Yamaha A520) and delivered through a loudspeaker (Infinity Primus P162, Stamford,

CT). All the training and testing behavioral performances were controlled and monitored through a custom-built MATLAB GUI. All trial events and behavioral responses were recorded and stored in the computer for further analysis and assessment of behavioral performance.

C. Training procedures for the three range-classification task

1. Basic paradigm

The present experiment employed a positive-reinforcement Go/Nogo paradigm similar to that described in a previous study (Yin *et al.*, 2010). Briefly, animals were trained to lick a waterspout as the correct behavioral response to a target sound and to refrain from licking the waterspout following presentation of a non-target sound. In detail, each training session began with initial delivery of a large drop of water (~ 0.5 ml) to initiate licking of the waterspout. After this initial drinking, animals were then trained to refrain from licking the waterspout for a minimum of 0.5 s in order to initiate a trial. They were trained to resume licking the waterspout within a 2.0 s response window after presentation of a target sound [*Go-trial*, Fig. 1(A)] for a water reward, or to refrain from licking the waterspout during the 2.0 s response window after hearing a reference sound [*Nogo-trial*, Fig. 1(B)]. Hits or correct Go-trial responses (resumed licking of the waterspout during the response window) were rewarded with a small drop of water (0.1–0.3 ml). Misses (on go-trials) were not rewarded. However, false alarms (incorrect licking during No-go trials) resulted in a 5–10 s timeout penalty, which was applied after the trial. The trials in which the animals licked in the early time window were discarded from the data analysis. The training session ended when the animal was no longer thirsty and did not lick the waterspout for three consecutive target trials.

2. Training procedures

After a 1 to 2 day habituation period, during which animals were familiarized with the testing box and learned to obtain water by licking the waterspout, the animals began training on the Three Range-Classification Task. Each trial consisted of the presentation of a single sound (0.5 s duration) drawn from one of the three defined ranges (Table I). Two ferrets (Group 1: Nile and Ganges) were trained on the task based on single feature dimensions. They were first trained on the task with a set of 15 tones and learned to perform the task based on tone frequency (Tone-task). After reaching behavioral criterion on the Tone-task, the two animals were then trained on a second task with a set of six amplitude-modulated (AM) noises and learned to perform the task based on the AM rate (AM-task). Two other ferrets (Group 2: Basil and Fennel) were trained on a combined features task with a set of six compound stimuli, consisting of different AM rates over different tone carriers (AM-Tone in Table I), and learned to perform the task based on the combination of frequency and AM rate (AM-Tone-task). During early stages of training in the Tone-task and AM-Tone-task, an intensity cue (for target sounds) was used to help shape animal responses to the

FIG. 1. Task paradigm. After the animal withheld licking from the waterspout (for at least 0.5 s) a trial began with the presentation of a single sound. (A) Go-trial: animal was trained to lick the waterspout within the defined 2.0 s response window when the sound was from the middle zone along the training feature dimensions. The animals were rewarded with a drop of water for a correct response (hit); (B) Nogo-trial: animal had to restrain from licking the waterspout over an equivalent duration (2 s) withdraw response window. A 5–10 s timeout penalty was added during the inter-trial-interval for an incorrect response (false alarm). (C) Example of a psychometric function obtained by sigmoid function fitting of the behavioral response rate over the stimulus parameter (frequency) across the boundary between the lower and middle frequency ranges. Two horizontal dashed lines indicate the range of the lick response rates and the two vertical dashed lines indicate the transition zone between the lower and middle ranges (see Sec. II for more details).

correct sound category (Go-sounds). During this training period, Nogo-sounds were presented at much lower sound intensity levels than the Go-sounds (up to an initial 50 dB relative attenuation of the Nogo sounds as indicated by the gray scale of filled markers in Figs. 2(A)–2(C)). The discrimination performance (discrimination rate or d' -prime) was computed for each training session, and was used to guide the adjustment of the attenuation level applied to Nogo-sounds. The intensity cues were gradually removed during training. In

contrast, no intensity cues were needed or used for training on the AM-task (as the animals had already learned the basic paradigm following Tone-task training).

3. Initial stimulus sets

The initial stimulus sets varied for different animals, as shown in Table I. However, in all cases, the stimuli from every set were divided into three regions along the relevant feature dimension, which were given behavioral meaning (one middle Go-region and two flanking Nogo-regions). For the Tone-task there were 15 tone frequencies (5 frequencies in each of the three defined ranges: Low, middle, and high). The absolute range of frequencies varied between animals. One animal (Nile) had small steps between adjacent tones [12% or ~ 2 semitone (st)] and hence spanned a fairly narrow frequency range (~ 2.3 octaves), whereas the other ferret (Ganges) had relatively larger steps between adjacent tones (41% or ~ 6 st) and spanned a much wider frequency range (~ 7 octaves). Neither the duration of training to reach criterion nor the level of final performance on the Tone tasks depended upon the initial stimulus set range for these two animals. The initial training set on the AM-Tone-task was sparser, and consisted of a total of six stimuli, two in each defined range. The frequency component in the AM-Tone-task (for Basil and Fennel) had step sizes of 100% (1 octave) and spanned a 5 octave range from 250 Hz to 8 kHz. The same six AM rates were used in both the AM-task and the AM-Tone-task for initial training, and covered a range from 4 to 60 Hz (see Table I).

D. Generalization test

After the animals learned the tasks with the initial stimulus sets, generalization tests were performed with a larger set of stimuli that had finer resolution and smaller step sizes (increments) between adjacent stimuli. The frequency increment for the generalization test was ~ 0.5 st (for Nile) and ~ 1.5 st (for Ganges) in the Tone-task, and was ~ 1 st for the frequency component in the AM-Tone-task. The generalization test on the temporal feature dimension (AM rate) had an increment of 1 Hz in both the AM-task and the AM-Tone-task. There were 18–19 equally sampled stimuli in each defined range, which made a total of 54–57 stimuli over the training range. The majority of the stimuli in the generalization tests were novel to the animals. The animals heard each of the new stimuli very few times during each generalization tests (no more than 4 times in one test session). Correct

TABLE I. Stimuli set for initial training.

Tasks	Animals	Low (Nogo)	Middle (Go)	High (Nogo)
Tone:	Ganges	125, 176, 249, 350, 494 (Hz)	697, 982, 1385, 1953, 2753 (Hz)	3882, 5474, 7719, 10 883, 15 345 (Hz)
	Nile	359, 402, 450, 504, 565 (Hz)	633, 709, 794, 889, 996 (Hz)	1115, 1549, 1399, 1566, 1754 (Hz)
AM:	Ganges	4, 15 (Hz)	26, 37 (Hz)	48, 59 (Hz)
	Nile	(carrier: white noise)	(carrier: white noise)	(carrier: white noise)
AM-Tone:	Basil	AM rate (carrier frequency):	AM rate (carrier frequency):	AM rate (carrier frequency):
	Fennel	4(250), 15(500) Hz	26(1000), 37(2000) Hz	48(4000), 59(8000) Hz

responses to a stimulus in the Go-range were rewarded during generalization tests. In each trial, the stimulus was chosen pseudo-randomly to make sure all stimuli were presented equally. To reduce possible training effects, the generalization test was conducted only once a week (1 out of 5 behavioral sessions/week) and animals were tested with the initial (small) trained stimulus set during the remaining four weekdays in order to maintain task performance. The response probability along the continuous axis of trained feature dimensions was fitted by a sigmoid function to define the stimulus boundaries between the Go and Nogo behavior.

E. Probe testing

The probe stimuli were chosen pseudo-randomly during the testing. In contrast to the generalization tests, animals could make a response freely to the probes, and there were no consequences for animals’ decision making (e.g., no reward, no timeout penalty irrespective of Go/Nogo choice). Probe stimuli constituted 10% of the total stimuli presented in a block of trials for a given task. In order to compare the relative salience across the two trained feature dimensions, the probe parameters were chosen so as to yield equal discriminability between the two feature dimensions (frequency and temporal).

1. Group 1

Multiple types of probe stimuli were tested during task performance in the two animals that had been trained on single feature-dimension tasks (Ganges and Nile in Table II). The probe stimuli were (1) AM-Tone, e.g., combinations of two learned feature dimensions (carrier frequency and AM rate) during performance of both the Tone-task and the AM-task; (2) tones (the learned frequency feature) when animals were engaged in the AM-task (the frequency feature was irrelevant in performing the task); or (3) AM-noises (the learned temporal feature) when animals were engaged in the Tone-task (the AM feature was irrelevant in performing the task).

2. Group 2

Similar probe stimuli were used in animals trained on the AM-Tone-task (Basil and Fennel in Table II). The training stimuli in AM-Tone-task were narrowband, and so to determine if there were any effects of this bandwidth choice, we also tested additional probe stimuli during the AM-Tone-task that varied in bandwidth. These probes were constructed of bandpassed noise (BPN) with bandwidths of 1, 3, 6, and

12 st. The center-frequency of the BPN probes were chosen at 250, 1000, and 8000 Hz, (the same as the Tone-probes, see Table II), for a total of 12 BPN probe stimuli (4 bandwidths × 3 frequencies). Two variants of these BPN probes were further tested: (1) BPNs that varied across frequency and bandwidth dimensions but did not have any AM; and (2) BPNs with AM rates at 4, 26, and 59 Hz [the same as the AM-probes (see Table II)].

F. Data analysis

1. Behavioral assessment

The timing of the first-lick (FL) after each sound presentation was recorded from each trial. All metrics used for assessing the behavioral performance were derived in relation to FL. The animal was considered to have made a *response* in a given trial if the FL fell within the response time window (0.2–2.2 s after sound onset). Otherwise, if the ferret did not make its first lick during the response window, the trial was marked as a *non-response*. A *hit* was defined as a *response* to a Go sound, and a *false alarm* was defined as a response to a Nogo sound. The hit rate (HR) and false alarm rate (FR) were computed in each training session. Two basic metrics were used in quantifying the behavioral performance based on HR and FR: (1) discrimination rate (DR), which was defined as: $DR = HR \cdot (1 - FR)$; and (2) d-prime, a measure based on signal detection theory and computed as: $d\text{-prime} = z(HR) - z(FR)$, where z denoted the computed $z\text{-score}$ of the HR and FR (Ahroon and Pastore, 1977). The criteria for significant behavioral performance in distinguishing Go and Nogo sound during training were defined as either $DR \geq 40\%$ or $d\text{-prime} \geq 1$. The criterion for task mastery was that animals achieved significant performance in three consecutive training sessions in the absence of any intensity cues for targets (i.e., all stimuli were presented at equal loudness). Animals continued training for a few weeks in order to consolidate their performance before they began the phase of generalization testing and probe testing.

2. Psychometric function and fitting

The psychometric functions of task performance were described by the behavioral *response probability* across stimulus feature dimensions (frequency or AM rate). We used the most common 4-parameter logistic (sigmoid) function to fit the psychometric function from low range to middle range to define the lower boundary, and also from middle range to high range to define the upper boundary, between the perceptual Go and Nogo ranges

TABLE II. The probe stimuli.

Animals	Task	Tone	AM-noise	AM-Tone
Ganges:	Tone-task	Na	4, 26, 59 (Hz)	AM rate (carrier frequency)
	AM-task	100, 784, 17210 (Hz)	Na	26 (100), 4 (784), 26 (784), 59 (784), 26 (17210) Hz
Nile:	Tone-task	Na	4, 26, 59 (Hz)	AM rate (carrier frequency)
	AM-task	359, 664, 1670 (Hz)	Na	26 (359), 4 (664), 26 (664), 59 (664), 26 (1670)
Basil:	AM-Tone:	250, 1000, 8000 (Hz)	4, 26, 59 (Hz)	AM rate (carrier frequency)
Fennel:				26 (250), 4 (1000), 59 (1000), 26 (8000)

$$y = a + \frac{b - a}{\left(1 + \frac{c}{x}\right)^d}.$$

Curve fitting was performed using MATLAB's "fminsearch" function, in which the searching started at $[a=0, b=1, c=0.5, d=5]$ for the lower boundary function fitting and at $[a=1, b=0, c=0.5, d=5]$ for the upper boundary function. Several useful measures could be derived from the fitted sigmoid curve [Fig. 1(C)]. The *inflection point* (c) of the sigmoid fitted curve [the point at which the function reaches the middle between lower plateau (a) and the upper plateau (b)] is taken as the perceptual boundary between the range categories. The *saturation point* was measured from the crossing at the upper plateau level of a tangent line through the *inflection point*, and the *threshold point* was computed from the crossing at the lower plateau level of the tangent line. The difference between the saturation and the threshold defined a *transition zone* between the perceptual Go and Nogo ranges. The *sensitivity* at the boundaries could be described by the slope at the inflection point that is derived from the parameter d .

3. Statistical testing

One-way analysis of variance (ANOVA) was performed to compare the behavioral responses to different stimuli, and hence determine if there was an equal response among the stimuli in each of the probe sets. If the animal had responded randomly to the probe stimuli within the Go- or Nogo-range, the measured DR would have varied between 0 and 0.25. So to test the significance of the discriminative performance among the probe stimuli or their feature components, we applied one-sample t -test for the hypothesis that "DR is greater than 0.25." If the p -value of the t -test was less than 0.05, the probe stimuli or their feature components are assigned a behavioral meaning (Go or Nogo). The discriminative performance among the different groups (e.g., task vs probe, or probes in different task) was tested with the Wilcoxon rank-sum test to determine the significance of the difference. Two-way ANOVA testing was used in assessing the effects of the task-demand and stimulus behavioral-meaning on the behavioral response in animals trained on single feature tasks, and the effect of the temporal (AM) and bandwidth features on behavioral response in animals trained on the AM-Tone-task.

G. Multiresolution representation of sound in an auditory cortical model

Animal performance was predicted using a multiresolution representation of the stimuli inspired by auditory cortical processing (Chi *et al.*, 2005; Yang *et al.*, 1992). All details of the model used are available in these papers, so we shall only briefly describe the highlights of the two stages of audio processing performed: (1) The first stage transforms the sound into an auditory spectrogram, and (2) the second stage performs a spectrotemporal analysis of the spectrogram modulations. The spectral analysis of the acoustic signal in the cochlea is modeled as a bank of 128 constant- Q asymmetric bandpass filters equally spaced on the logarithmic

frequency scale spanning 5.2 octaves. The filter outputs are then transduced into inner hair cell potentials via a high-pass and low-pass operation. The resulting auditory nerve signals undergo further spectral sharpening via a lateral inhibitory network. Finally, another transformation is performed by a midbrain model, leading to additional loss in phase locking by using short term integration with a time constant of 4 ms, resulting in a time frequency representation called the auditory spectrogram. The second stage analyzes the spectrotemporal content of the auditory spectrogram using a bank of modulation-selective filters centered at each frequency along the tonotopic axis, mimicking neurophysiological receptive fields. This step corresponds to a two-dimensional (Frequency $\times$ time) affine wavelet transform, with a spectrotemporal mother wavelet, defined as a Gabor-shape in frequency and exponential in time. Each filter h is tuned ($Q=1$) to a specific rate (ω in Hz) of temporal modulations and a specific scale of spectral modulations (Ω in cycles/octave), and a direction of movement ($\pm$). For an input spectrogram $z(t, f)$, the response of each cortical filter then is given by

$$r^\pm(t, f; \omega, \Omega; \vartheta, \varphi) = z(t, f) * t, f h^\pm(t, f; \omega, \Omega; \vartheta, \varphi),$$

where $*t, f$ denotes convolution in time and frequency and ϑ and Φ are the characteristic phases of the cortical filters

FIG. 2. Timecourse of training and task acquisition. The plots show the timecourse of the discriminative performance across each daily training session from individual animals (indicated by different markers). An attenuation cue to enhance the salience of the Go-stimuli was applied during early training sessions (as indicated by the level of the gray scale filled on the marker). The horizontal dashed line in each plot indicates the behavioral criterion for a significant discriminative performance of the task. The vertical lines mark the session number at which the animal reached criterion for task performance (three consecutive daily sessions above criterion at 0 dB attenuation—i.e., equal loudness for all sounds). (A) Two Group 1 animals trained on the Tone-task; (B) two Group 1 animals trained on the AM-task; (C) two Group 2 animals trained on the AM-Tone task.

which determine the degree of asymmetry in the time and frequency axes, respectively. The MATLAB software that performed all this analysis is available for download from our lab website (<http://www.isr.umd.edu/Labs/NSL/Software.htm>).

Predictions of animal performance were made by computing the cortical output for any probe signal, and then comparing it to the representation of all training signals. The Euclidian distance between the two representations was computed and normalized by the sum of all distances between the probe and the training stimuli. For a given probe, the behavioral response rate (Go-response) was predicted by

$$R' = a + be^{(cx)},$$

where x denotes the normalized cortical distance between a probe and the trained “Go”-stimulus; R' denotes the predicted rate of a Go response; a , b , c are the coefficients which could be determined by best fitting the function in least-squares sense (MATLAB function *lsqcurvefit* started at [0 100 1]).

III. RESULTS AND DISCUSSIONS

A. Task acquisition

All four animals from Groups 1 and 2 successfully learned the three range-classification task based on the selected features, although the number of sessions to reach

behavioral criterion varied significantly across animals from 22 to 82 sessions. The progress of task acquisition for the different animals is shown in Fig. 2. Each marker represents the discrimination performance (d-prime) for a given training session, and the filled gray scale indicates the attenuation levels on the Nogo-sounds during the early training. The horizontal lines indicate the significance level of the performance criterion (d-prime at 1.0). The varied number of sessions to learn the task was related to the parameters used during the training, such as the range of intensity cues, the attended feature dimension, the size of the stimulus set, the increment size between adjacent stimuli in the set, the previous training history, and also the individual differences between the animals. The two animals trained on the AM-Tone-task [both traces in Fig. 2(C)] and one animal on the Tone-task [the circle in Fig. 2(A)], received a similar range of intensity cues (40–50 dB) during training, and required a similar number of training sessions (22–23 sessions) to reach the task criterion. The ferret (Nile) who learned the Tone-task [the square in Fig. 2(A)] started with smaller initial intensity cues (~20 dB) and smaller increments between the stimuli (2 st), and required 47 sessions to reach behavioral criterion. Learning the AM-task, in general, required more training sessions than learning the Tone-task (51 vs 23 in Ganges, 82 vs 47 in Nile). The presence of intensity cues in the Tone-task, and the fact that all the ferrets learned the Tone-task before learning the AM-task might be important factors influencing the time-course of learning the AM-task.

(A) Tone-task (Ganges,n=5)

(B) Tone-task (Nile,n=9)

(C) AM-task (Ganges,n=4)

(D) AM-Tone-task (Basil,n=10)

FIG. 3. Task generalization and perceptual boundaries. The plots from (A)–(D) are examples of response profiles across stimulus parameters during generalization tests from: the Tone-task with large (A) or small (B) frequency separation, the AM-task (C), and the AM-Tone task (D). The data are represented as mean plus standard error from all generalization test sessions. The vertical dashed lines indicate the corresponding frequencies/AM rates during the initial training. The thick curves are the psychometric functions obtained by sigmoid fitting of the behavioral response vs stimulus parameters around the boundaries between the low and middle ranges (black), and the middle and high ranges (gray). The vertical black lines indicate the perceptual boundaries between different ranges.

However, it is also possible that more training is needed to learn tasks involving temporal features and future behavioral studies will be needed to clarify this question.

B. Generalization test

After animals learned all required tasks and established stable performance, generalization tests were conducted across the trained feature dimensions with a finer step size than the original training set stimuli. Figures 3(A)–3(D) show examples of the mean percentage of response across the test stimulus feature dimensions during the generalization test in different task: Tone-task with larger increments [Fig. 3(A)], Tone-task with smaller increments [Fig. 3(B)], AM-task [Fig. 3(C)], and AM-Tone-task [Fig. 3(D)]. There were significantly higher response rates to the sounds in the middle (Go) range than to either low or high (Nogo) ranges during the generalization tests in all tasks. The gradient of response rate from low to middle range or from middle to high range was fitted nicely with a sigmoid (logistic) function, forming the lower and upper plateaus corresponding to the perceptual Go- and Nogo-ranges and a steep transition between the perceptual ranges. There was a relatively small variation in the response rate within each range, possibly because the ferrets treated each range as a defined and perceptually distinguished acoustic category.

C. Probe tests on group 1 animals: The task-engagement and feature interaction

Group 1 animals were trained on single feature dimension tasks and switched freely from the Tone-task to the AM-task task in blocks or even on a trial-by-trial basis. This allowed us to probe the effects of task engagement on the retrieval of behavioral meaning of learned features (see Table II). Specifically, we tested how the animals responded during a session performing one task to probes introduced from the other task with a different learned feature. That is, how did the animals interpret AM-noises as probes while engaged in the Tone-task, and vice versa? Figure 4 shows the test results from one animal (Ganges). The animal responded equally to the tones whether they were presented in the context of the Tone-task or AM-task [gray versus black bars in Figs. 4(A) and 4(B)]. Thus, the response to the probe stimuli showed a strong correlation with their behavioral meaning, e.g., Go or “Nogo” [Fig. 4(C)], and there were no statistical differences in discriminative performance between the probes in a different task context and the same stimuli presented during performance of the appropriate task [Wilcoxon rank sum test: $p = 0.8293$ (Tones), $p = 0.1039$ (AM-noises)]. Nor was there a statistical difference between the reaction times to the probe and the same sounds in either task (Fig. S1¹).

A second question concerns how the two learned features interact in the retrieval of their behavioral meaning from a compound probe: AM-Tone (combination of frequency and AM rate). The AM-Tone probes included one correlated combination and four anti-correlated feature combinations, in which the probe stimuli were constructed with Go in one feature and Nogo for another feature of the same

FIG. 4. Probe testing in Group 1 animals: Single features. (A) The average behavioral response to Tones and AM-noise probes during the Tone-task. (B) The average behavioral response to AM-noises and Tone probes during the AM-task. (C) The comparison of the discriminative performance among the task stimuli (black bar) and among the same stimuli played as probe stimuli (gray bar) during the alternate task. All data were represented as mean plus standard error (the same for the boxplots in Figs. 5–7) (p -values are from Wilcoxon rank sum test).

stimulus (see Table II). If the correct retrieval of an AM-Tone stimulus was dependent on whether the feature was engaged in the current task, we would predict a different response profile to AM-Tone probes during performance of a Tone-task or an AM-task. For example, when an animal was engaged in the Tone-task, we might expect the responses to the compound probes to correlate with the spectral feature (the carrier frequency) of the probes, or vice versa when the animal was engaged in the AM-task. Our broad finding, however, was that in both animals, regardless of the ongoing task, the responses to the probe stimuli were positively correlated primarily with the behavioral meaning of the frequency feature, and not with the AM features of the compound stimuli [Figs. 5(A1-2) and 5(B1-2)]. Nevertheless, there were some weak but significant task effects (suppressed responses to the AM-Tone probes) when one animal (Ganges) engaged in the AM-task [black vs gray bars in Fig. 5(A3)] [Fig. 5(A3), two-way ANOVA, $p = 0.0385$]. Further analysis of the reaction times to the AM-Tone probes with carrier frequency in Go range (Fig. S2¹), revealed a significantly delayed response in Ganges to the AM-Tone probe compared to the same probe sound during performance of the Tone-task ($p = 0.0321$) [Fig.

(A) AM-Tone probe (Ganges)

(B) AM-Tone probe (Nile)

FIG. 5. Probe testing in Group 1 animals: Combination AM-Tone probes. (A1), and (B1) The average behavioral response to Tones and AM-Tone probes during the Tone-task. (A2) and (B2) The average behavioral response to AM-noises and AM-Tone probes during the AM-task. (A3) and (B3) The comparison of the behavioral response to individual AM-tone probes between the animals engaged in the Tone-task (black bar) and in AM-task (gray bar). There is a weak but significant task effect (two-way ANOVA testing task $\times$ stimuli) on the behavioral response to AM-Tones probes observed in one animal (A3), but not in another animal (B3), indicating an individual difference between the two animals. (A4) and (B4) The discriminative performance between task stimuli (black bar) and AM-Tone probes according to the carrier frequencies (dark gray bar) or the AM rate (light gray bar). All data were represented as mean plus standard error (ns = no significant difference, $*p < 0.5$, $**p < 0.01$, $***p < 0.001$). The p -values are computed from a Wilcoxon rank-sum test between groups or one-sample t -test).

S2(A1)¹]; but no such difference was found in the second animal [Nile; Fig. S2(B1)¹]. This might explain the difference in the discrimination performance among the AM-tone probes between the two animals, indicating significant variation among individuals.

To summarize, there were significant discriminative performances in both animals along the spectral dimension [gray bars in Figs. 5(A4) and 5(B4)], but not the temporal dimension [lighter gray bars in Fig. 5(A4) and 5(B4)], of the probe stimuli. The animals, therefore, seem to have interpreted the combination AM-Tone probes primarily as Tones, and proceeded to associate them with their meaning in the Tone-task. It is likely that the narrow bandwidth of the carrier (a pure tone) may have been the critical cue that directed the animals' behavior.

D. Probe tests on group 2 animals: What did the animal learn from AM-Tones?

Group 2 animals were trained on an AM-Tone-task, where the two-feature combination stimuli had correlated combinations of spectral (frequency) and temporal (AM rate) cues (Table II). The animals, in fact, could have

performed this task with at least three different strategies—either by attending to both of the correlated features, or to either one of the feature dimensions. To test the ferrets' task strategy and attentional focus, we conducted tests with three types of probes (Table II): (a) Pure-Tone (the carrier frequency of AM-Tone), (b) AM-noise (white noise carrier modulated at different AM rates), and (c) AM-Tone (four anti-correlated combinations of AM rates and frequencies).

The two animals demonstrated similar performance patterns across the different probe stimuli, i.e., significant discrimination for both (1) the Pure-Tone probes [frequency feature only, one-way ANOVA: $p = 0.0002$ (Basil) and $p = 0.0042$ (Fennel)], and also (2) the AM-Tone probes [one-way ANOVA: $p = 0.0002$ (Basil) and $p = 2.126 \times 10^{-9}$ (Fennel)]. However, we found only (3) marginal or no significant discrimination to the AM-noise probe [temporal feature only, one-way ANOVA: $p = 0.5147$ (Basil) and $p = 0.0371$ (Fennel)] [Figs. 6(A1) and 6(B1)]. Overall, the discriminative performance among all probes was driven more strongly by the carrier frequency than by the temporal feature (AM rate), as seen by the larger spectral (versus temporal) discriminative performance in Figs. 6(A2) and 6(B2) in both animals. This result is consistent with the performance of

(A) Probes in AM-Tone-task (Basil)

(B) Probes in AM-Tone-task (Fennel)

FIG. 6. Probe testing in Group 2 animals. (A1) and (B1) The average behavioral response to AM-Tones (task stimuli) and the different probes during performance of the Tone-task. (A2) and (B2) The discriminative performance among the task stimuli (black bar) and the different probes are based on the carrier frequencies (dark gray bar) or the AM rate (light gray bar). All data were represented as mean plus standard error [ns = no significant difference, $*p < 0.5$, $**p < 0.01$, $***p < 0.001$, p -values were from one-way ANOVA in (A1) and (B1), and Wilcoxon rank-sum test between groups or on-sample t -test in (A2) and (B2)].

Group 1 animals to the AM-tone probes (Fig. 5), in which they responded primarily to the behavioral meaning of the carrier frequency rather than its AM rate. However, it is also evident that responses to individual AM-Tone probes (the anti-correlated combinations) were sometimes modulated by the AM rates, e.g., the significant dependence on AM-rate for carrier frequency 1000 Hz. There were no systematic differences in the reaction times between the task stimuli and the probes (Fig. S3¹).

The response to AM-Tone probes demonstrates a lack of behavioral meaning for the temporal feature dimension of the compound stimuli [lighter gray bars in Figs. 6(A2) and 6(B2)]. However, the discriminative performance among all probe stimuli [gray bar in Figs. 6(A2) and 6(B2)] was significantly less than performance in the task [(black bars in Figs. 6(A2) and 6(B2))]. This indicates that although there is relatively little behavioral meaning carried in the temporal dimension compared to the spectral dimension, still, the envelope fluctuations in the training stimulus set form an implicit stimulus context that might be a contributing factor in stimulus memory and therefore may play a role in shaping task performance.

E. Probe testing on compound-feature task: Exploring the effects of the bandwidth

Tone probes alone did *not* elicit maximal performance in animals trained on the AM-Tone task [Figs. 6(A) and 6(B)].

One possible explanation is that the lack of envelope fluctuations (compared to the AM-tones), made these stimuli somewhat different in timbre and that this dissimilarity hence reduced the animals' performance level. This conjecture was tested by a series of probes that varied from pure-tone, through narrowband noises (BPN) with increasing bandwidth (BW) and centered at the appropriate frequencies, and also modulated at different AM rates. We found that the response to the BPN probes during the task depended both on the center frequency and bandwidth of the probe stimuli [gray lines in Figs. 7(A1) and 7(B1)]. The maximum response was elicited when the bandwidth was approximately 1 st in both spectral and AM probes. It then decreased with increasing or decreasing bandwidth [Figs. 7(A2) and 7(B2)]; ANOVA2: BW factor, $p < 0.001$ in both animals. These results suggest that animal performance diminishes the further away the stimulus BW was from the original training stimuli. At each BW, the best response always occurred when the center frequency or the AM rate was in the Go range [dark and light gray lines in Figs. 7(A1) and 7(B1)]. These results are consistent with the hypothesis that the envelope fluctuations in the original training set of the AM-Tone stimuli comprised an implicit context memory that constrained discriminative performance, since the spectral range of the envelope fluctuation from the BPN at 1 st is closely matched by the fluctuation range within the original training stimulus set (see Fig. S4¹).

(A) Bandwidth effects (Basil)

(B) Bandwidth effects (Fennel)

FIG. 7. Bandwidth effects in Group 2 animals during AM-Tone-task. (A1) and (B1) Shows the average behavioral response to task stimuli (AM-Tones, the black line) and to the probes (the two gray lines) with different bandwidths during performing AM-Tone task. Two sets of probes were tested: the BPN probes (the darker gray line) and the amplitude-modulated BPN probes (the lighter gray line). The arrowheads in the plot indicate the corresponding task stimuli which have the same carrier frequency (but differ in bandwidth) or AM rate with the probes. (A2) and (B2) The discriminative performance is computed among the task stimuli (the dashed lines) and the probe sounds at different bandwidth with (the light gray solid line) or without (the darker gray solid line) envelope AM. All data were represented as mean plus standard error. (ns = no significant difference, $*p < 0.5$, $**p < 0.01$, $***p < 0.001$, p -values were based one-way ANOVA test).

IV. PREDICTIONS OF A PATTERN CLASSIFIER BASED ON CORTICAL REPRESENTATION OF SOUND

In an attempt to explain the entire complex pattern of behavioral results in both groups of animals, we computed the predictions made by a simple classifier which consisted of the following stages: (1) sound is represented by a basic multiresolution spectrotemporal model of cortical processing (Chi *et al.*, 2005), briefly reviewed in Sec. II; and (2) the animals retain in memory this cortical representation of the training stimuli (as templates) together with their meaning (or labels as Go or Nogo); and finally, (3) a novel probe stimulus is interpreted according to the label of the nearest stored template, but weighted in proportion to its distance from it. We shall refer to this classifier as the cortical-classifier. It is essentially the well-known nearest-neighbor-classifier (Duda *et al.*, 2001), but using the multiresolution cortical representation of the acoustic stimuli for measuring and comparing distances.

The performance predictions computed from this classifier account reasonably well for the average performance of the animals for all probes in Group 1 and 2 animals. Especially remarkable is the classifier's prediction of seemingly arbitrary changes in performance across different probes (Fig. 8), such as that exhibited by both Group 2 animals for probes (1000 Hz, 4 Hz) compared to (1000 Hz, 59 Hz) in Fig. 6. This pattern of results can be explained in a manner consistent with the rest of the findings as follows: (1) Both probe tones have a carrier frequency of 1000 Hz, and hence are closest in percept to the third training stimulus (carrier frequency = 1000 Hz, and AM rate = 26 Hz). This is due to the high saliency of the spectral cue (carrier frequency) relative to temporal cue (AM rate). Since the third training stimulus is a go stimulus, this also explains the large overall response rate. (2) Next, the cortical model predicts that the 59 Hz probe resembles (or is closer to) the 26 Hz

(third training stimulus) than is the 4 Hz probe. The reason is that the 4 Hz rate is almost twice as far from the 26 Hz rate as the 59 Hz rate (2.7 octaves compared to 1.2 octaves). The cortical model rate filters are about 1 octave in BW, and hence the 4 Hz cortical patterns diverge from the 26 Hz patterns significantly more than the 59 Hz patterns. The combination of these two factors explains this result.

Other predictions that are consistent with the measured performance of Group 2 animals include:

- (1) The weak but significant tuning of the pure tone probes (relative to AM-tone training stimuli): This is explained in the cortical model by the fact that tones match the carrier frequencies of the training AM-tones, but differ from them by the envelope AM which somewhat diminishes the performance.
- (2) The weak tuning to the AM noise probes: Here the match between the probes and training stimuli is on the temporal rates dimension. However, the two sets of stimuli differ considerably in their carriers (tones versus white-noise), which therefore significantly diminishes the retrieval performance.
- (3) The maximum response to BPN (1 st) and gradual decrease with increasing bandwidths for all other BPN probes (Fig. 7): This is explained by the fact that, as the envelope modulation rates increase, the spectrum of the AM-tone training stimuli effectively slightly broaden (approximately <100 Hz), and hence the spectrum of the probes of bandwidth 1 st matches best that of the training stimuli, compared to that of the pure tones, or the progressively broader probes. Performance therefore is predicted to deteriorate progressively as the distance between the training and probe stimuli increases.

V. GENERAL DISCUSSION

Two groups of ferrets were trained on auditory LTM tasks and learned to retrieve the meaning of sounds with spectral and temporal cues. The two ferrets in Group 1 learned to classify tones based on their frequency (Tone-task) or amplitude modulation rate (AM-task). Animals could generalize their training to perform on novel stimuli along the trained feature dimensions, indicating that they had formed clear "range" classes during training. Moreover, animals could also switch rapidly between the two learned tasks on a trial-by-trial basis, easily classifying along either one feature dimension or the other without impairing their behavioral performance. This task-switching was clearly facilitated by immediately and implicitly recognizing the task context by the distinctive bandwidths of the carriers in each stimulus set (spectrally narrow in the Tone-task vs broadband noise in the AM-task), that were learned during training. Ferrets continued to rely on this implicit context cue when they responded to the compound AM-Tone probes, in that the spectrally narrowband carrier cued them to respond primarily to the spectral feature (frequency) rather than to the temporal feature (AM rate). Thus, the use of implicit context cues (stimulus set characteristics) in

FIG. 8. Predicted response rate vs the actual response rate among the variety of probes. The bar plots the predicted GO response rate based on the cortical distance between the probes and the Go-training stimulus, which overlay with the actual response rates to those probes from two Group 2 animals [the lines, the same data from Figs. 6(A1) and 6(B1)]. The predicted values are highly correlated with the actual response rates from the two animals [$R^2 = 0.9481$ (Basil), 0.7960 (Fennel)].

association with the selected feature-dimension in the task was revealed by the behavior of the animals in Group 1.

In contrast to the training of the ferrets in Group 1, the two ferrets in Group 2 learned to classify AM-tone stimuli based on their correlated feature combination (AM-Tone task). And while either the spectral or the temporal feature alone was sufficient to define the three acoustic zones, the results from a variety of probe tests, nevertheless, indicated that the animals preferred to respond to variations in the spectral cue (carrier frequency) during the task. Their performance, however, was also somewhat affected by cues along the temporal dimension (AM rate), as when responding to uncorrelated AM-Tones probes (Fig. 6). In addition, the animals also exhibited discriminative performance to the changing BW of the probes, performing best when the noise BW was comparable to that seen in the training stimuli (1 st).

To summarize, this complex pattern of results is consistent with the hypothesis that (1) the animals weighted the spectral dimension cues relatively highly compared to the AM rate (a form of primacy of frequency over rate); but (2) that *both* the spectral and the temporal dimensions (among others such as bandwidth) were attended to and exploited to perform the tasks; and (3) that, as a simple rule, the more a stimulus as a whole deviated from the training stimuli, the greater was the decrease in the animal's ability to retrieve its meaning correctly.

A. Feature based auditory long-term memory

Recent studies of auditory pattern memory and recognition have reported conflicting findings from a variety of species. For example, experiments with monkeys and dogs have raised the possibility that these animals are unable to store the representation of complex acoustic stimuli in long-term recognition memory (Kowalska *et al.*, 2001; Fritz *et al.*, 2005). By contrast, ethological studies have shown that birds (Lambrechts and Dhondt, 1995), seals and sea lions (Insley *et al.*, 2000; Charrier *et al.*, 2001; Pitcher *et al.*, 2010; Insley and Holt, 2012), and monkeys and apes (Cheney and Seyfarth, 1980; Rendall *et al.*, 1996; Ceugniet and Izumi, 2004; Candiotti *et al.*, 2013; Keenan *et al.*, 2016) all have LTMs of individual conspecific voices. Monkeys have also been shown to recognize conspecific predator alarm calls, as well as alarm calls from other species of monkeys and birds (Zuberbuhler, 2000). Finally, laboratory studies have also demonstrated the existence of an auditory LTM for simple acoustic stimuli in some animal species such as rats (Njegovan *et al.*, 1993, 1995), starlings (Zokoll *et al.*, 2007, 2008a), and songbirds (Weisman *et al.*, 1998, 2004).

Ferrets in our experiments demonstrated a clear auditory LTM. They were able to recall stimulus class membership and sort acoustic stimuli into three ranges based on not only the frequency feature dimension (Tone-task) but also on the temporal feature dimension (AM-task). In addition, ferrets also learned a different set of classes of complex acoustic stimuli (AM-Tones) and could sort these stimuli into three zones (AM-Tone-task), and were quite sensitive to other acoustic properties associated with the stimuli such as the bandwidth. However, a detailed examination of the results

suggests that the animals did not weight all features equally, preferring instead the spectral dimension when available. This result suggests that auditory representations and memory (in humans and animals) may be organized along multiple separate feature dimensions, making it possible to select and weight the most task-relevant dimension for memory, be it pitch, bandwidth, loudness, or timbre (Deutsch, 1970; Demany and Semal, 2008; Semal and Demany, 1991; Krumhansl and Iverson, 1992; Clement *et al.*, 1999). Aside from psychoacoustic experiments, there is also support for this view from studies of mismatch negativity (MMN). The MMN elicited by different feature deviants from the standard has different neural sources (location), indicating separate storage sites for basic perceptual acoustic attributes, such as stimulus frequency, intensity, duration, and location (Giard *et al.*, 1995; Deacon *et al.*, 1998). Such feature-based memory store is also supported by evidence, that when a deviant tone differs from the standard in both frequency and duration (termed a double deviant), the MMN is elicited by a frequency feature but not the duration feature (Czigler and Winkler, 1996). These results suggest that independent feature-based processing may be a key aspect of auditory sensory memory (Nousak *et al.*, 1996; Caclin *et al.*, 2006).

However, in contrast, other studies (Stilp *et al.*, 2010, 2011, 2012, 2016) have shown that two highly correlated features of complex sounds (they used attack/decay amplitude envelopes and spectral shape) may collapse into a single perceptual dimension that enhances efficient coding. While their model nicely explains their results on encoding of co-varying acoustic stimuli in human subjects after very brief, short-term passive, or active exposure to these sounds, under our quite different experimental conditions, with long-term training over months in an animal model, our behavioral evidence from the current ferret study argues for a more generalized hypothesis in which the correlated features do not collapse into a single perceptual dimension, but rather are treated as relatively independent templates that are memorized and used in subsequent recognition. Their model would predict that for our two ferrets (Basil and Fennel) trained on correlated features, there should have been poor responses to probe stimuli that deviated from the single (derived) correlated dimension (probes such as the unmodulated 1000 Hz stimulus, or a 1000 Hz stimulus modulated at 4 or 59 Hz). Instead, as can be seen in Fig. 6, for both animals, these two stimuli elicited significantly higher responses than would have been predicted from the Stilp-Kluender model. As mentioned above, a key difference between the two studies (in addition to species) is the effect of time course and duration of training—we note that many of the adaptive effects described by Stilp and Kluender only last a short time. This raises a question as to whether collapse into a single stimulus dimension is more likely during rapid adaptation to co-varying stimulus statistics in a short-term paradigm, whereas long-term auditory memory may also be encoded by independent weighting of acoustic dimensions or in templates for each feature dimension or specific templates for each acoustic exemplar. Further human studies to contrast and investigate these two possibilities (efficient coding of co-varying features in a single perceptual dimension, or

encoding of dual co-varying feature dimensions with relatively independent templates or multiple exemplars) are important to conduct.

B. Representation of learned acoustic features during retrieval from auditory LTM

Clearly the percept of any unified auditory object may bind together multiple features, causing interdependency in their detection by human listeners (Joseph *et al.*, 2015). Nevertheless, auditory features may still intrinsically vary in their effectiveness in the representation of such auditory objects in memory. For instance, in the study by Joseph *et al.*, memory recall was found to be more accurate along the spectral dimension than along the temporal dimension (AM rate), and concluded that detection along the spectral dimension alone produced comparable performance to the “object condition,” and that therefore for this particular dimension, holding an object or a feature in mind drew equally on memory resources. These are indeed highly relevant findings for our study. Thus, to assert that auditory representations are feature-based does not specify how features are weighted in the representation, nor whether these preferences are intrinsic, or are modulated by top-down attention and task-relevance.

Such attention-driven shifting of feature preference is evident, for example, in Group 1 ferrets, which were trained on both the Tone-task and AM-task and could attend to either spectral or temporal dimensions to perform the tasks. These ferrets could readily switch their attentional focus between the two dimensions when faced with a compound probe, based simply on an implicit bandwidth cue and regardless of their overall behavioral (task) context (Fig. 5). However, since ferrets were clearly able to utilize both spectral and temporal cues in different task contexts, we asked what general principles one might be able to derive from these findings regarding their use of other cues in future tasks?

The cortical classifier predicted well the pattern of the performance exhibited by all ferrets. This nearest-neighbor-classifier may at first glance suggest that the animals simply memorize all training stimuli and then compare and select the closest one. This approach theoretically works very well (Duda *et al.*, 2001) if the feature space is relatively small and appropriately chosen, and if the training samples are finite. Furthermore, the cortical feature representation of sound appears to do an adequate job of weighting more heavily the spectral dimension (the tonotopic axis) while still maintaining some dependence on the temporal (rate) and bandwidth (scale) dimensions that are known to be fundamental dimensions of auditory cortical processing (Schreiner *et al.*, 2000). The origin for this relatively enhanced sensitivity to the spectral (tonotopic) compared to the temporal (rate) dimension in the cortical model stems simply from the higher resolution of the former. Thus, small changes in the carrier frequency of the stimuli induce substantial translations of the response pattern along the spectral dimension largely because of the narrow bandwidths of the spectral filters (about 0.1 octave; Chi *et al.*, 2005). By contrast, the rate

filters (which extract and represent the AM rate modulations) have wider bandwidths (1 octave; Chi *et al.*, 2005), and hence translations along this dimension have to be substantially larger to induce the same detectable disparities. In the same manner, one may predict the sensitivity to changes in bandwidth (the scale axis of the cortical model), or to other features to be included in the model (such as loudness, pitch, and spatial location) based on the relative “resolution” of these dimensions, which are usually estimated from independent psychoacoustic and physiological measurements as was done for the axes of the current model.

Another exemplar-based recognition model [exemplar-based random walk (EBRW) model] that provides an account of multidimensional choice probabilities and category formation (that is indeed very similar to ours, measuring nearest-neighbor distances to stored templates of the training samples) has been developed by Nosofsky and colleagues over the past 30 years (Nosofsky *et al.*, 2011, 2014). His model proposes dynamic exemplar-based retrieval mechanisms that would result in the emergence of a familiarity-based evidence-accumulation process, accounting for the time course of categorization and recognition decision-making and reaction times. Given their fundamental formal compatibility, it would be possible to integrate our recognition model (with its cortical representation) with his EBRW model.

However, the nearest-neighbor-classifier is but one of a wide range of related classifiers, which can take into account more neighbors (k -neighbor-classifiers), with additional weighting and editing of distances (Samworth, 2012). It is quite likely therefore that these, and other linear or nonlinear classifiers (Duda *et al.*, 2001), can do well to account for our data provided that the cortical representations are utilized. Consequently, replacing the training samples with “frequency range” categories or even more broad categories about the task context (in Group 1 animals) may well be a good alternative, provided again that the category is represented in the appropriate feature space.

VI. CONCLUSIONS

Ferrets can be trained on an auditory LTM task to classify sounds over different ranges along spectral or temporal dimensions. Although it is possible to explain these results in terms of the formation of sound categories, by Occam’s razor, we favor a simpler, more parsimonious alternative explanation that uses a nearest-neighbor analysis and relies upon storage of trained template stimuli. According to this view, when faced with novel probes, the ferrets associate probe stimuli with the meaning of the nearest-neighbor training stimulus in their LTM. Our results show that their classification weights the spectral features of the sound carrier more heavily, but is still influenced by the contextual information contained in the temporal modulations and the bandwidth of the carrier. A cortical multi-resolution representation of sound largely accounts for the performance of the animals in different tasks. Future studies of auditory LTMs will explore the relative behavioral importance of

the many other acoustic features of sound such as intensity, timbre, and spatial location.

ACKNOWLEDGMENTS

We thank the undergraduates, Kayla Kahn, Eileen Chang, Adam LaFleur, and Sam Burgess, for their valuable contributions to this behavioral study through daily assistance in ferret training. We are also grateful for support from NIH Grant No. R01 DC005779.

¹See supplementary material at <http://dx.doi.org/10.1121/1.4968395> for the analysis of response latencies to different sets of probe and the spectrum of the BPN envelopes at different bandwidths.

- Ahroon, W. A., Jr., and Pastore, R. E. (1977). "Procedures for computing d' and β ," *Behavior Res. Meth. Instrument.* **9**(6), 533–537.
- Caclin, A., Brattico, E., Tervaniemi, M., Naatanen, R., Morlet, D., Giard, M., and McAdams, S. (2006). "Separate neural processing of timbre dimensions in auditory sensory memory," *J. Cognitive Neurosci.* **18**(12), 1959–1972.
- Candiotti, A., Zuberbuhler, K., and Lemasson, A. (2013). "Voice discrimination in four primates," *Behav. Process.* **99**, 67–72.
- Ceugniet, M., and Izumi, A. (2004). "Vocal individual discrimination in Japanese macaques," *Primates* **45**, 119–128.
- Charrier, I., Mathevon, N., and Jouventin, P. (2001). "Mother's voice recognition by seal pups," *Nature* **412**, 873.
- Cheney, D. L., and Seyfarth, R. M. (1980). "Vocal recognition in free-ranging Vervet monkeys," *Animal Behav.* **28**, 362–367.
- Chi, T., Ru, P., and Shamma, S. A. (2005). "Multiresolution spectrotemporal analysis of complex sounds," *J. Acoust. Soc. Am.* **118**(2), 887–906.
- Clement, S., Demany, L., and Semal, C. (1999). "Memory for pitch versus memory for loudness," *J. Acoust. Soc. Am.* **106**, 2805–2811.
- Czigler, I., and Winkler, I. (1996). "Preattentive auditory changes detection relies on unitary sensory memory," *Neuroreport* **7**(17017), 2413–2417.
- David, S. V., Fritz, J. B., and Shamma, S. A. (2012). "Stimulus valence and task-relevance control rapid plasticity in primary auditory cortex," *Proc. Natl. Acad. Sci. U.S.A.* **109**, 2144–2149.
- Deacon, D., Noursak, J. M., Piloti, M., and Ritter, W. (1998). "Automatic change detection: Does the auditory system use representation of individual stimulus features or gestalts?," *Psychophysiol.* **35**, 413–419.
- Demany, L., and Semal, C. (2008). "The role of memory in auditory perception," in *Auditory Perception of Sound Sources*, Vol. 29 in *Springer Handbook of Auditory Research Series*, edited by W. A. Yost and R. R. Fay (Springer, New York), pp. 77–113.
- Deutsch, D. (1970). "Tones and numbers: Specificity of interference in immediate memory," *Science* **168**, 1604–1605.
- Duda, R. O., Hart, P. E., and Stork, D. G. (2001). *Pattern Classification*, 2nd ed. (John Wiley & Sons, New York), 680 pp.
- Fritz, J., Mishkin, M., and Saunders, R. C. (2005). "In search of an auditory engram," *Proc. Natl. Acad. Sci. U.S.A.* **102**, 9359–9364.
- Fritz, J. B., Shamma, S. A., Elhilali, M., and Klein, D. (2003). "Rapid task-dependent plasticity of receptive fields in primary auditory cortex," *Nat. Neurosci.* **6**, 1216–1223.
- Giard, M. H., Lavikainen, J., Reinikainen, K., Perrin, F., Bertrand, O. M., Perier, J., and Naatanen, R. (1995). "Separate representation of stimulus frequency, intensity and duration in auditory sensory memory: An event-related potential and dipole model analysis," *J. Cognitive Neurosci.* **7**, 133–143.
- <http://www.isr.umd.edu/Labs/NSL/Software.htm>. The site includes Matlab Toolbox for modeling auditory cortical processing, which was developed by Neural Systems Laboratory (Last viewed on November 16, 2016).
- Innsley, S. J. (2000). "Long-term vocal recognition in the northern fur seal," *Nature* **406**, 404–405.
- Innsley, S. J., and Holt, M. M. (2012). "Do male northern elephant seals recognize individuals or merely relative dominance rank?," *J. Acoust. Soc. Am.* **131**(1), EL35–EL41.
- Joseph, S., Kumar, S., Husain, M., and Griffiths, T. D. (2015). "Auditory working memory for objects vs. features," *Front Neurosci.* **9**, 13.
- Keenan, S., Mathevon, N., Stevens, J. M., Guery, J. P., Zuberbuhler, K., and Leverro, F. (2016). "Enduring voice recognition in bonobos," *Sci. Rep.* **6**, 22046.
- Kowalska, D. M., Kusmierek, P., Kosmal, A., and Mishkin, M. (2001). "Neither perirhinal/entorhinal nor hippocampal lesions impair short-term auditory recognition memory in dogs," *Neurosci.* **104**, 965–978.
- Krumhansl, C. I., and Iverson, P. (1992). "Perceptual interaction between musical pitch and timbre," *J. Exp. Psychol. Human Percept. Perform.* **18**(3), 739–751.
- Lambrechts, M. M., and Dhondt, A. A. (1995). "Individual voice recognition in birds," *Curr. Ornithol.* **12**, 115–139.
- Njegovan, M., Ito, S., Mewhort, D., and Weisman, R. (1995). "Classification of frequencies into ranges by songbirds and humans," *J. Exp. Psychol.* **21**(1), 33–42.
- Njegovan, M., Weisman, R., Ito, S., and Mewhort, D. (1993). "How grouping improving the categorization of frequency in song birds and humans and why song birds do it better," *J. Canadian Acoust. Assoc.* **21**(3), 87–88.
- Nosofsky, R. M., Cox, G. E., Cao, R., and Shiffrin, R. M. (2014). "An exemplar-familiarity model predicts short-term and long-term probe recognition across diverse forms of memory search," *J. Exp. Psychol. Learn., Memory Cognition* **40**(6), 1524–1539.
- Nosofsky, R. M., Little, D. R., Donkin, C., and Fific, M. (2011). "Short-term memory scanning viewed as exemplar-based categorization," *Psychol. Rev.* **118**(2), 280–315.
- Noursak, J. K., Deacon, D., Ritter, W., and Vaughan, H. G., Jr. (1996). "Storage and comparison of information in transient auditory memory," *Cognitive Brain Res.* **4**, 305–317.
- Pitcher, B. J., Harcourt, R. G., and Charrier, I. (2010). "The memory remains: Long-term vocal recognition in Australian sea lions," *Animal Cognition* **13**(5), 771–776.
- Polley, D., Read, H. L., Storace, D. A., and Merzenich, M. M. (2007). "Multiparametric auditory receptive field organization across five cortical fields in the albino rat," *J. Neurophysiol.* **97**(5), 3621–3638.
- Rendall, D., Rodman, P. S., and Emong, R. E. (1996). "Vocal recognition of individuals and kin in free-ranging rhesus monkeys," *Animal Behavior* **51**, 1007–1015.
- Samworth, R. J. (2012). "Optimal weighted nearest neighbor classifiers," *Ann. Stat.* **40**(5), 2733–2763.
- Scott, B. H., Mishkin, M., and Yin, P. (2012). "Monkeys have a limited form of short-term memory in auditory," *Proc. Natl. Acad. Sci. U.S.A.* **109**, 12237–12241.
- Scott, B. H., Mishkin, M., and Yin, P. (2013). "Effect of acoustic similarity on short-term auditory memory in the monkey," *Hear. Res.* **298**, 36–48.
- Schreiner, C. E., Read, H. L., and Sutter, M. L. (2000). "Modular organization of frequency integration in primary auditory cortex," *Annu. Rev. Neurosci.* **23**, 501–529.
- Semal, C., and Demany, L. (1991). "Dissociation of pitch from timbre in auditory short-term memory," *J. Acoust. Soc. Am.* **89**, 2404–2410.
- Stilp, C. E., and Kluender, K. R. (2011). "Non-isomorphism in efficient coding of complex sound properties," *J. Acoust. Soc. Am.* **130**(5), EL352–EL357.
- Stilp, C. E., and Kluender, K. R. (2012). "Efficient coding and statistically optimal weighting of covariance among acoustic attributes in novel sounds," *PLoS One* **7**(1), e30845.
- Stilp, C. E., and Kluender, K. R. (2016). "Stimulus statistics change sounds from near-indiscernible to hyperdiscernible," *PLoS One* **11**(8), e0161001.
- Stilp, C. E., Rogers, T. T., and Kluender, K. R. (2010). "Rapid efficient coding of correlated complex acoustic properties," *Proc. Natl. Acad. Sci. U.S.A.* **107**(50), 21914–21919.
- Weisman, R. G., Njegovan, M. G., Sturdy, C. B., Phillmore, L., Coyle, J., and Mewhort, D. (1998). "Frequency-range discriminations: Special and general abilities in zebra finches (*Taeniopygia guttata*) and humans (*Homo sapiens*)," *J. Comp. Psychol.* **112**(3), 244–258.
- Weisman, R. G., Njegovan, M. G., Williams, M. T., Cohen, J. S., and Sturdy, C. B. (2004). "A behavior analysis of absolute pitch: Sex, experience, and species," *Behav. Processes* **66**(3), 289–307.
- Yang, X. W., Wang, K., and Shamma, S. A. (1992). "Auditory representation of acoustic signal," *IEEE Trans. Inf. Theory* **38**(2), 824–839.
- Yin, P., Fritz, J. B., and Shamma, S. A. (2010). "Do ferrets perceive relative pitch?," *J. Acoust. Soc. Am.* **127**(3), 1673–1680.
- Yin, P., Fritz, J. B., and Shamma, S. A. (2014). "Rapid spectrotemporal plasticity in primary auditory cortex during behavior," *J. Neurosci.* **34**(12), 4396–4408.

- Zokoll, M. A., Klump, G. M., and Langemann, U. (2007). "Auditory short-term memory persistence for tonal signals in a songbird," *J. Acoust. Soc. Am.* **121**, 2842–2851.
- Zokoll, M. A., Klump, G. M., and Langemann, U. (2008b). "Auditory memory for temporal characteristics of sound," *J. Comp. Physiol. A* **194**, 457–467.
- Zokoll, M. A., Naue, N., Herrmann, C. S., and Langemann, U. (2008a). "Auditory memory: A comparison between humans and starlings," *Brain Res.* **1220**, 33–46.
- Zuberbuhler, K. (2000). "Interspecies semantic communication in two forest primates," *Proc. R. Soc. London B* **267**, 713–718.