European Journal of Neuroscience Proposed section: Special issue Auditory Cortex 2014

Meaning in the avian auditory cortex: Neural representation of communication calls

Julie E Elie and Frédéric E Theunissen

Helen Wills Neuroscience Institute and Department of Psychology, University of California Berkeley, Berkeley, CA, USA

Corresponding author :
Julie E Elie
Helen Wills Neuroscience Institute and Department of Psychology, University of California Berkeley, Berkeley, CA, USA
3210 Tolman Hall, University of California, Berkeley, CA-94720
julie.elie@berkeley.edu and julie.elie@gmail.com

Running Title: Meaning in the avian auditory cortex

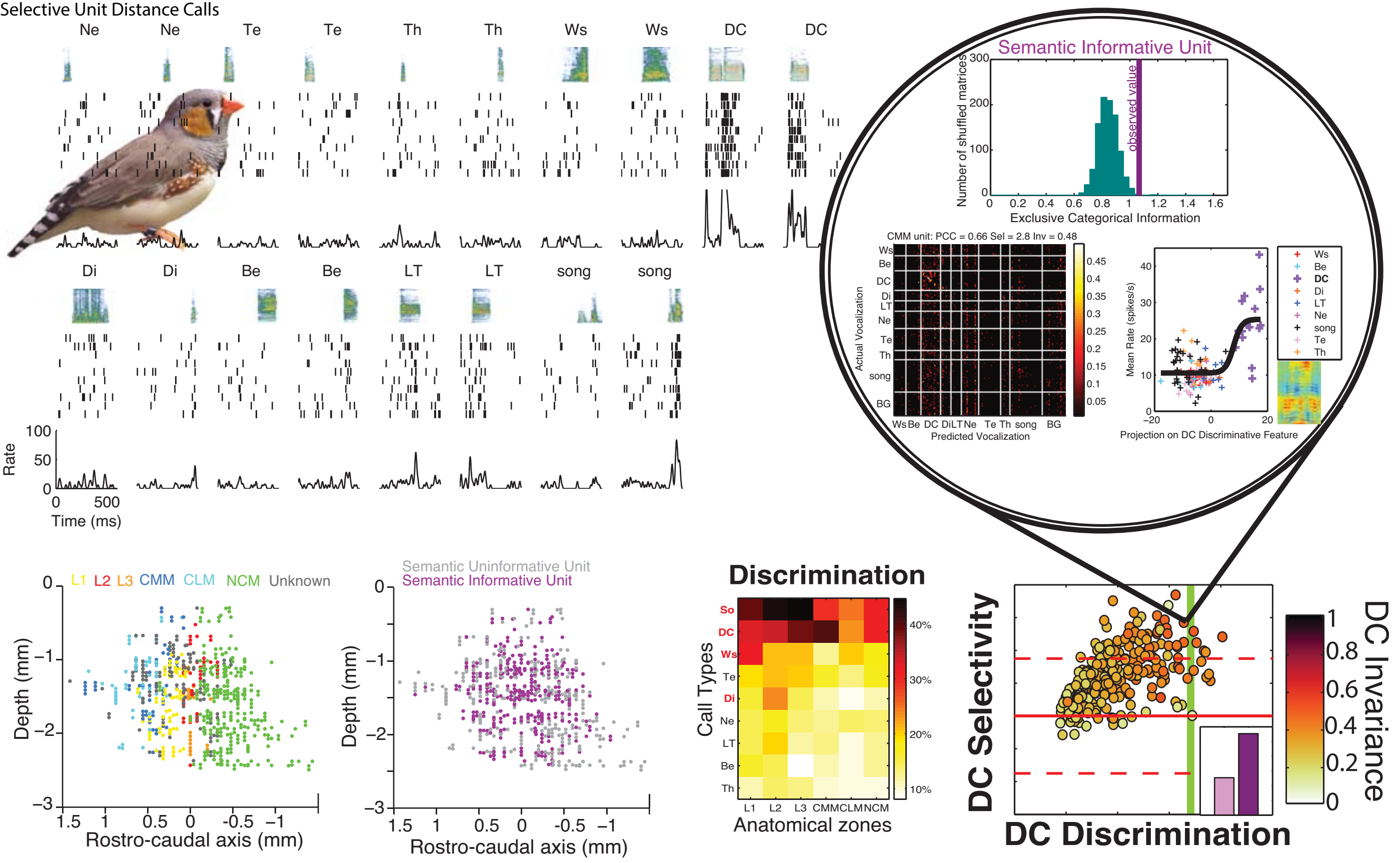
Total number of pages: 59
Total number of figures: 9 + 4 supplementary
Total number of tables: 2
Total number of equations:12

Total number of words: in the whole manuscript: 14781; in the Abstract: 234; in the Introduction: 520.

Keywords: zebra finch, semantics, categorization, selectivity, invariance.

GRAPHICAL ABSTRACT

Using a very large library of communication calls we investigated how semantics is represented in the avian auditory cortex. Almost half of the single units exhibit responses that can be used to discriminate semantic categories and information theoretic calculations show that single units provide about 10% of the maximum achievable information. Neurons encode semantics with various degrees of selectivity and invariance that are the result of non-linear transformations on acoustical features.


Selective Unit Distance Calls
Ne
Ne
Te
Te
Th
Th
Ws
Ws
DC
DC
Di
Di
Be
Be
LT
LT
song
song
Rate
100
50
0
0
500
Time (ms)
Semantic Informative Unit
Number of shuffled matrices
300
200
100
0
observed value
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
Exclusive Categorical Information
CMM unit: PCC = 0.66 Sel = 2.8 Inv = 0.48
Actual Vocalization
Ws
Be
DC
Di
LT
Ne
Te
Th
song
BG
Ws Be DC Di LT Ne Te Th song BG
Predicted Vocalization
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
Mean Rate (spikes/s)
40
30
20
10
0
−20
0
20
Projection on DC Discriminative Feature
Ws
Be
DC
Di
LT
Ne
song
Te
Th
L1 L2 L3 CMM CLM NCM Unknown
Depth (mm)
0
−1
−2
−3
1.5
1
0.5
0
-0.5
−1
Rostro-caudal axis (mm)
Semantic Uninformative Unit
Semantic Informative Unit
Depth (mm)
0
−1
−2
−3
1.5
1
0.5
0
-0.5
−1
Rostro-caudal axis (mm)
Discrimination
Call Types
So
DC
Ws
Te
Di
Ne
LT
Be
Th
L1
L2
L3
CMM
CLM
NCM
Anatomical zones
40%
30%
20%
10%
DC Selectivity
DC Discrimination
DC Invariance
1
0.8
0.6
0.4
0.2
0

Abstract

Understanding how the brain extracts the behavioral meaning carried by specific vocalization types that can be emitted by various vocalizers and in different conditions is a central question in auditory research. This semantic categorization is a fundamental process required for acoustic communication and presupposes discriminative and invariance properties of the auditory system for conspecific vocalizations. Songbirds have been used extensively to study vocal learning, but the communicative function of all their vocalizations and their neural representation has yet to be examined. In our research, we first generated a library containing almost the entire zebra finch vocal repertoire and organized communication calls along 9 different categories based on their behavioral meaning. We then investigated the neural representations of these *semantic* categories in the primary and secondary auditory areas of 6 anesthetized zebra finches. To analyze how single units encode these call categories, we described neural responses in terms of their discrimination, selectivity and invariance properties. Quantitative measures for these neural properties were obtained using an optimal decoder based both on spike counts and spike patterns. Information theoretic metrics show that almost half of the single units encode semantic information. Neurons achieve higher discrimination of these semantic categories by being more selective and more invariant. These results demonstrate that computations necessary for semantic categorization of meaningful vocalizations are already present in the auditory cortex and emphasize the value of a neuro-ethological approach to understand vocal communication.

Introduction

Although vocal communication is essential for the survival of many animal species, the neurophysiological basis of the perception of intra-specific communication signals is still not well understood. Vocalizations used by animals are rich signals that contain information about the vocalizer identity and spatial location and some “meaning” that refers to the emotional status, the intent of the vocalizer or even referential information such as a particular type of food in advertising calls or a particular type of predator in alarm calls (Marler, 2004a; Seyfarth & Cheney, 2010; Manser, 2013). Depending on the type of information a listener is attending to (e.g. meaning: presence of a danger), the variability of the acoustic signals that relates to the identity of the vocalizer (voice recognition), the production variance between renditions (Gentner, 2004) or even the transmission variability (spatial location and propagation effects; (Mouterde *et al.*, 2014)) might be of less importance and could be ignored by the listener (Tsunada & Cohen, 2014). Identifying the neural basis of the discrimination of acoustic communication signals and the tolerance or invariance of responses to vocalizations that share the same meaning but differ acoustically is challenging. To study such natural categorization, one must use a model species for which the *semantics* of vocalizations can be easily derived from the observation of social behaviors. Here, we use *semantics* to describe call categories obtained from the meaning of vocalizations that is inferred by the behavioral contexts in which vocalizations are emitted. Also, a perfect model species should be easily reared in laboratory conditions while still producing most of the vocalizations in its repertoire during social interactions with peers (Bennur *et al.*, 2013). Many studies on auditory categorization have used artificial categories of sounds as stimuli, after extensive training in discrimination tasks (Jeanne *et al.*, 2011; Tsunada *et al.*, 2011; Meliza & Margoliash, 2012; Tsunada *et al.*, 2012). Although this research has revealed the critical role of secondary auditory areas in the representation of categories, the generalization to the processing of conspecific communication signals is questionable since intensive operant training might influence the neural processes of perception and could be different from social learning (Gentner & Margoliash, 2003; Bieszczad & Weinberger, 2010; David *et al.*, 2012; Bennur *et al.*, 2013). The use of conspecific vocalizations in primate neurophysiological studies has shown that spatial and semantic information are processed by two different streams (respectively the dorsal and ventral streams) in the auditory cortex of monkeys (Rauschecker & Tian, 2000; Tian *et al.*, 2001; Romanski *et al.*, 2005; Rauschecker & Scott, 2009; Romanski & Averbeck, 2009; Bizley & Cohen, 2013). This extensive work has begun to reveal the categorization properties at different stages of the ventral pathway: from the categorization of spectro-temporal features in the core region of the auditory cortex, to the categorization of abstract features such as call semantics in the superior temporal gyrus (STG) and the ventro-lateral prefrontal cortex (vlPFC) (Gifford *et al*., 2005b; Cohen *et al*., 2006; Cohen *et al*., 2009; Tsunada *et al*., 2012; Tsunada & Cohen, 2014). However, these studies contain clear limitations. The captive macaques were never reared in an environment that would enable them to hear and learn the usage of their own conspecific vocalizations while socially interacting with peers (but see (Gifford *et al*., 2003) for some discrimination of food calls by captive macaques). Also, the limited vocalization bank (Hauser, 1998) did not allow an extensive investigation of the invariance of neural representations to the voice characteristics of different vocalizers. In the present study, we used a comprehensive vocalization library to investigate the neural representation of communication calls in a social songbird species, the zebra finch. We employed a decoding model of the spiking activity of single neurons to explore where and how semantic information is encoded in the avian auditory cortex. We quantified the discrimination and selectivity properties of neurons to meaningful categories as well as the invariance of neural responses to vocalizer identity.

METHODS

Animals

Four male and 2 female adult zebra finches (*Taeniopygia guttata*) from the Theunissen Lab colony were used for the electrophysiological experiments. The birds were bred and raised in family cages until they reached adulthood, and then maintained in uni-sex groups. Although birds could only freely interact with their cage-mates, all cages were in the same room allowing for visual and acoustical interactions between all birds in the colony.

Twenty-three birds (8 adult males, 7 adult females, 4 female chicks and 4 male chicks) were used as subjects for the acoustic recordings of zebra finches vocalizations. Nine of the adults (5 males and 4 females) and all of the chicks were from the Theunissen Lab colony while six adults (3 males, 3 females) were borrowed from the Bentley Lab colony

(University of California, Berkeley) for the time of the recording period (2-3 months). We used these two origins to increase the inter-individual variability of vocalizations. During the period of audio recordings, adult birds were housed in groups of 4 to 6 birds (2 to 3 pairs) and each group was acoustically and visually isolated from the other birds. Chicks were housed in a family cage with their parents and siblings.

All birds were given seeds, water, grid and nest material *ad libitum* and were supplemented with eggs, lettuce and bath once a week. All animal procedures were approved by the Animal Care and Use Committee of the University of California Berkeley and were in accordance with the NIH guidelines regarding the care and use of animals for experimental procedures.

Stimuli

Vocalizations used as stimuli during neurophysiological experiments were recorded from 15 adult birds and 8 chicks (20-30 days old). Adults were recorded while freely interacting in mixed-sex groups in a cage (L = 56 cm, H = 36 cm, D = 41 cm) placed in a sound proof booth (Med Associates Inc, VT, USA). During each daily recording session (147 sessions of 60 to 90 minutes), a handy digital recorder (Zoom H4N Handy Recorder, Samson; recording parameters: stereo, 44100 Hz) was placed 20 cm above the top of the cage while an observer monitored the birds' behavior hidden behind a blind. Chicks were also recorded with the same audio recording device while interacting with their parents in a cage (L = 56 cm, H = 36 cm, D = 41 cm) placed in a sound proof booth (Acoustic Systems, MSR West, Louisville, CO, USA). To elicit begging calls, chicks were isolated from their parents for 30 minutes to 1 hour before recording. Based on the observer notes, individual vocalizations from each bird were manually extracted from these acoustic recordings and annotated with the identity and sex of the emitter and the social context of emission. The vocalization bank obtained contains 486 vocalizations (see Table 1).

Following Zann's classification of vocalization categories (Zann, 1996), we used the acoustical signatures and behavioral context to classify the vocalization into 7 semantic categories in adults and 2 in chicks. In adults we found:

- Song: multi-syllabic vocalization (duration in our dataset: 1424±983 ms; mean ± sd) emitted only by males either in a courtship context (directed song) or outside of a courtship context (undirected song).
- Distance call: loud and long (duration in our dataset: 169±49 ms) monosyllabic vocalization used by zebra finches to maintain acoustic contact when they can't see each other.
- Tet call: soft and short (duration in our dataset: 81±16 ms) monosyllabic vocalization emitted by zebra finches at each hopping movement to maintain acoustic contact with the nearest individuals.
- Nest call: soft and short (duration in our dataset: 95±75 ms) monosyllabic vocalization emitted around the nest by zebra finches that are looking for a nest or are constructing a nest. This category grouped together the Kackle and Ark calls described by Zann (Zann, 1996) since these two categories formed a continuum in our recordings and were hard to dissociate.
- Wsst call: long (503±499 ms in our dataset) noisy broad band monosyllabic or polysyllabic vocalization emitted by a zebra finch when it aggressively supplants a cage-mate.
- Distress call: long (452±377 ms in our dataset), loud and high-pitched monosyllabic or polysyllabic vocalization emitted by a zebra finch when escaping from an aggressive cage-mate.
- Thuk call: soft short (53±13 ms in our dataset) monosyllabic vocalization emitted by birds when there is an imminent danger but they are reluctant to flee.

For chicks, we distinguished 2 call or semantic categories:

- Long Tonal call: loud and long (184±63ms) monosyllabic vocalization that chicks emit when they are separated from their siblings or parents. The Long Tonal call is the precursor of the adult Distance call.

- Begging call: loud and long (382±289 ms in our dataset) monosyllabic call emitted in bouts when the bird is actively begging for food to one of its parent (lowering its head and turning its open beak in direction to the parent beak).

These 9 call categories encompass almost all call types found in the complete repertoire of the wild Zebra finch (Zann, 1996). We did not include Whine calls and Stack Calls. Whine calls are also produced during nesting and pair-bonding behavior and although we recorded many Whines in our domestic zebra finches we did not capture a large enough number of examples from each of our subjects to include them in our neurophysiological analyses. Stack calls are produced in wild zebra finches at takeoff and are described as being intermediate between Tets and Distance calls. We did not record or were not able to distinguish Stack calls in our domesticated birds. Durations of vocalizations were obtained in two steps. First, we calculated the RMS intensity of identified sound periods in the waveform. The sound periods were defined as any sequence of non-null values in the sound pressure waveform longer than 20 ms. Second, the actual boundaries of the vocalizations were obtained by finding the window in the sound period where the rectified signal was above 35% of the RMS intensity.

For the neurophysiological experiments, a new subset of the vocalization bank was used at each electrophysiological recording site (n=25). This subset was made from a representative subset of vocalizations from the repertoire of 10 individuals: three adult females, three adult males, two female chicks and two male chicks. The identity of the individuals was randomized between sites except for one male, one female, one male chick and one female chick; vocalizations from these four birds were broadcast at every single electrophysiological recording site. For each site, a subset of the vocalizations from each bird was obtained by random selection of 3 Wsst calls, 3 Distance calls, 3 Distress calls, 3 Nest calls, 3 Songs, 3 Tet calls, 3 Thuk calls, 3 Begging sequences, 3 Long Tonal calls (Fig Supp1). When birds had 3 or fewer calls in a given call category, all calls were used. The average number of stimuli per vocalization category played back at each electrophysiological site is given in Table 1. Our recording protocol was designed to obtain 10 trials per stimulus at each recording site but this number of trials varied slightly as we sometimes lost units before the end of a recording session and sometimes ran additional trials; on average each single vocalization was played 10±0.22 times (mean ± sd). Vocalizations were band-pass filtered between 250Hz and 12kHz to remove any low or high frequency noise. This range of frequencies is larger than the hearing range of the zebra finch (Amin *et al*., 2007). The sound pressure waveforms of the stimuli were normalized within each category to remove the intra category variability while preserving the natural average differences of sound levels between vocalization categories. A 2ms cosine ramp was applied at the beginning and at the end of each stimulus to create short fade in and fade out. Finally sounds were down sampled to 24414.0625 Hz to match the sampling rate of the processor used to broadcast the stimuli during the neurophysiological recordings (TDT System III, Tucker Davis Technologies Inc, FL, USA,).

Surgery

Twenty-four hours prior to the actual recording of neurons, the subject was fasted for an hour, deeply anesthetized with isoflurane (2L/min to initiate anesthesia and 0.8-1.6L/min to maintain state) and immobilized in a stereotaxic system so as to maintain its head with an angle of 50° with the vertical. After sub-cutaneous injection of 150 µL lidocaïne, its scalp was removed and a homemade head holder was glued to the outer layer of the skull using dental cement (Dentsply Caulk). The subject was housed alone in a cage for recovery until acute recording. On the morning of electrophysiological recordings the bird was fasted for 1 hour prior to anesthesia with urethane 20% (75 µL total in 3 injections in the pectoral muscles every half hour). The subject was placed back in the stereotaxic system using the head holder so its ears were free of any device. For the whole surgery procedure and recording session, the body temperature was maintained between 39 and 40ºC with a heating pad. Two rectangular openings of 2 mm long and 0.5 mm large, centered at 0.95 mm lateral in the left hemisphere, 0.5mm lateral in the right hemisphere and 1.25 mm rostral to the Y sinus, were created in both layers of the skull and the Dura to enable electrode penetration. An electrode array of two rows of 8 tungsten electrodes (TDT, diameter 33µm, length 4mm, electrode spacing 250µm, row spacing 500µm) was lowered in each hemisphere. To target all 32 electrodes to the avian auditory cortex, electrodes in the left hemisphere were inserted from the left with a 15° angle to the vertical in the coronal plane and electrodes in the right hemisphere were inserted from the caudal part of the bird with a 17° angle to the vertical in the sagittal plane. Note that for one of the subjects, only one electrode array in the left hemisphere was used. Before penetration, electrodes were coated with DiI powder (D3911, Invitrogen, OR, USA) to enable tracking in histological slices.

Electrophysiology

Extra-cellular electrophysiological recordings were performed in a sound-attenuated chamber (Acoustic Systems, MSR West, Louisville, CO, USA), using custom code written in TDT software language and TDT hardware (TDT System III). Sounds were broadcasted in a random order using an RX8 processor (TDT System III, sample frequency 24414.0625 Hz) connected to a speaker (PCxt352, Blaupunkt, IL, USA) facing the bird at approximately 40cm. The sound level was calibrated on song stimuli to obtain playbacks at 75dB SPL measured at the bird's location using a sound meter (Digital Sound Level Meter, RadioShack). Neural responses were recorded using the signal of two (5 subjects) or one (1 subject) 16-electrode arrays, band-pass filtered between 300Hz and 5kHz and collected by an RZ5-2 processor (TDT System III, sample frequency 24414.0625 Hz). Spike arrival times and spike shapes of multiple units were obtained by voltage threshold. The level of the threshold was set automatically by the TDT software using the variance of the voltage trace in absence of any stimuli. Electrodes were progressively lowered and neural responses were collected as soon as auditory responses to song, white noise, Distance call or limited modulation noise (Hsu *et al*., 2004b) could be identified on half of the electrodes in each hemisphere (the stimuli used to identify auditory neurons were different from the stimuli used in the analysis). Several recording sites were randomly selected by progressively deepening the penetration of the electrodes and ensuring at least 100 µm between two sites. On average 4.2±2 sites (mean ± sd) were recorded per bird and per hemisphere at a depth ranging from 400 µm to 2550 µm.

Histology

After the last recording site, the subject was euthanized by overdose of isoflurane and transcardially perfused with 20 mL PBS then 50-100mL paraformadehyle 4% pH=7.4. After dissection, the brain was sunk in paraformaldehyde 4% overnight to achieve good fixation, then cryoprotected in 30% sucrose-PBS. Once the brain showed the same density as the sucrose solution (usually after 48h), it was progressively frozen using liquid Nitrogen and stored in a freezer (-20ºC). Coronal slices of 20µm obtained with a cryostat were then alternatively stained with Nissl staining or simply mounted in Fluoroshield medium (F-6057, Fluoroshield with DAPI, Sigma-Aldrich). The slides were visualized on a light microscope (Zeiss AxioImager) and the images were digitized using a high-resolution digital CCD camera (Hamamatsu Orca 03). While Fluoroshield slices were used to localize electrode tracks, Nissl stained slices were used to identify the position of the 6 auditory areas investigated here: the three regions of Field L (L1, L2 and L3), 2 regions of Mesopallium Caudale (CM): Mesopallium Caudomediale (CMM) and Mesopallium Caudolaterale (CLM); and Nidopallium Caudomediale (NCM). By aligning pictures, we were able to anatomically localize most of the recording sites (672 out of 914 single units) and calculate the approximate coordinates of these sites. Since we could not localize the Y-sinus on slices, we used the position of the Lamina Pallio-Subpallialis (LPS) peak as the reference point for the rostro-caudal axis in all subjects. The surface of the brain and the midline were the reference for respectively the dorsal-ventral axis and the medial-lateral axis. The approximate coordinates of units were used to build 3-D reconstructions of all single units positions in an hypothetic brain, with a custom algorithm written in Matlab (Mathworks, Cambridge, MA).

Data Analysis

Sound analysis of the stimuli

To interpret the results of the neurophysiological recordings, we first analyzed the relationship between acoustical features and semantic categories using three measures. First we quantified the similarity between vocalizations within and across categories by cross-correlation analyses of their spectrograms. Second we used linear discriminant analysis (LDA) on spectrograms to quantify the discriminability of semantic categories in a configuration that maximized differences between all categories. Third, we used logistic regression classifiers on the spectrograms to quantify the discriminability of semantic categories in a one-vs-all-others configuration. Since different stimulus ensembles were used at each recording site, we calculated the within category correlations and performed the LDA for each ensemble. In this manner, we could directly compare acoustical properties of an ensemble of vocalizations to the neural responses to this same stimulus ensemble. For all three acoustical analyses, we used an invertible spectrographic representation (Singh & Theunissen, 2003) instead of extracting specific features such as, for example the spectral mean. Using an invertible spectrogram has the advantage of having the potential to capture any information bearing acoustical feature with the disadvantage of requiring many parameters for describing sounds, which demands additional approaches to prevent over-fitting (see below). The spectrogram of each vocalization was obtained using Gaussian windows of temporal bandwidth of ~ 3ms (corresponding to a spectral bandwidth of ~ 50 Hz) as measured by the "standard deviation" parameter of the Gaussian. The total length of the temporal window was taken to have 6 standard deviation and is therefore ~ 18 ms. All spectrograms had 234 frequency bands between 0 and 12 kHz and a sampling rate of ~1 kHz. For the within category cross-correlation analysis, we used the same 600 ms analysis frames that was used to estimate peristimulus time histograms (PSTH). This 600 ms frame required 611 points in time. For the LDA and logistic regression we used 200 ms analysis frames requiring 201 points in time. A shorter time window was required for the LDA because we wanted to isolate each syllable of polysyllabic vocalizations.

Stimulus cross-correlation

Before calculating cross-correlation in the spectrograms of stimulus pairs, the two vocalizations were aligned using the delay that gave the maximum cross-correlation value between the temporal amplitude envelopes of the stimuli (obtained from the spectrogram by summing the amplitudes across all frequency bands at each time point). The correlation between the two stimuli was then estimated by the correlation coefficient calculated between the overlapping zones of the aligned spectrograms. Fig Supp2A shows a matrix of correlation values obtained between the stimuli of one of the sets of vocalizations used during neurophysiological recordings. For each set of vocalizations used as stimuli, the average correlations within each category and between each category and all the others were calculated. Fig Supp2B gives the mean and standard deviation of these values across vocalization sets.

Semantic category discriminability: LDA and Logistic Regression

For these analyses, the 600ms stimuli were first cut into individual elements that were all of the same length (200 ms) and time aligned. This step ensured that vocalizations comprised by several individual elements (Wsst, Distress, Begging calls and Songs) would be separated into single sound elements. To isolate single sound elements, we estimated the sequence of maxima and minima in the temporal amplitude envelope of each stimulus. The amplitude envelope was estimated by full rectification of the sound pressure waveform followed by low-pass filtering below 20 Hz. Sound segments were defined as all the points above 10% of the maximum overall amplitude of the stimulus and, conversely, silence was defined as all the points below 10%. The maximum of each sound segment and the minimum of each silence segment were found and used to cut the vocalization bouts into individual elements. Sound segments shorter than 30 ms were ignored while those longer than 30 ms were aligned by finding the *mean time* and centering this time value at 100 ms (i.e. middle of the 200 ms frame). The mean time is obtained by treating the amplitude envelope as a density function of time (Cohen, 1995), and corresponds to the center of mass of the amplitude envelope. Sounds longer than 100 ms on either side of the mean time were truncated while those shorter than 100ms on either side of the mean time were padded with zeros. After sectioning, the spectrograms of sound elements were calculated as explained above. To reduce the number of dimensions of the spectrographic representation and prevent over-fitting of the discriminant algorithms, we performed a Principal Component Analysis (PCA) on the spectrograms of the sound elements. The number of Principal Components (PCs) used in the LDA was determined both by examining the cumulative fraction of the variance explained, and by performing the LDA with varying numbers of PCs (from 10 to 300 PCs). Because the performance of the classification that was achieved in cross-validated data sets peaked at 50 PCs, we used the first 50 PC coefficients as parameters in the LDA. Moreover, since the cumulative fraction of the variance explained by 50 PCs was approximately 85% of the total variance, we are confident that our database of vocalizations was sufficiently large to use LDA directly on the spectrograms (and not on a small number of acoustical parameters such as is often done in bio-acoustical research). Throughout the article, this method of discrimination on spectrograms is called PCLDAS (Principal Component Linear Discriminant Analysis on Spectrograms).

To further demonstrate the selectivity and invariance properties of neural responses, we also performed a series of logistic regression analyses, one for each semantic category. The goal of these analyses was to find the unique linear combination of acoustical features that would allow one to separate one type of vocalization from all the others. The inputs to the logistic regression were taken to be the coordinates of each call in the subspace defined by the significant discriminant functions obtained in the LDA.

Neural data analysis

Sorting multi-units to select auditory single units.

688 multi-units (2 x 16 x 18 + 16 x 7= 688) were recorded using the protocol described above. These units were sorted into single units based on spike shape (spike sorting) and twice sorted for the quality of their neural responses to sounds: before and after spike sorting. This process yielded 914 single auditory units.

To identify units responsive to sounds, we quantified the reliability and strength of the neural activity in response to auditory stimuli by estimating the coherence between a single spike train (R) and the actual time-varying mean response (A). This value of coherence γ_{AR}^{2} can be derived from the coherence between the peristimulus time histogram (PSTH) obtained from half of the trials and the PSTH obtained from the other half (Hsu *et al.*, 2004b):

$$\gamma_{AR}^{2} = \left[1 - \frac{M}{2} \times \left(1 - \sqrt{\frac{1}{\gamma_{\bar{R}_{1,M/2}\bar{R}_{2,M/2}}^{2}}} \right) \right]^{-1}$$

with M the total number of trials (presentations of the stimuli) and $\gamma_{\bar{R}_{1,M/2}\bar{R}_{2,M/2}}^{2}$ the coherence between the two PSTHs calculated on half of the trials. The coherence between two responses is a function of frequency (ω). An overall quantifier of the reliability and strength of the neural response was obtained by integrating over all frequencies (Hsu *et al.*, 2004a):

$$I_{AR} = -\int_{0}^{\infty} \log_2 \left[1 - \gamma_{AR}^{2}(\omega) \right] . d\omega$$

Here I_{AR} is expressed in information units (bits per second) and is an estimate of the mutual information between R and A responses (Borst & Theunissen, 1999). The thresholds used on I_{AR} to consider a unit as responsive to sounds was 3 bits/s for multi-units, yielding 658/688 auditory units before spike sorting, and 2.3 bits/s for spike-sorted units. These thresholds were chosen so that none of the units below that threshold had a spiking rate significantly increased by any stimulus.

Spike sorting of multi-units was performed using a semi-automatic custom program written in Matlab that used both un-supervised (k-means) and supervised clustering algorithms (Random Forest). In a first stage, templates for single spike shapes were chosen by the user using a GUI and exploratory cluster analysis for each multi-unit. For this process, several random groups of 200 spikes were successively clustered into 6 groups using k-means algorithm. The k-means clustering used the coefficients of a PCA performed on all the spikes' shapes of that multi-unit. The user could then manually select and assign the groups of spikes that were clearly belonging to the different single units constituting the multi-unit. In a second stage, these templates were used to train a Random Forest that used the PCA coefficients and additional spike parameters: the max and min amplitude and the peak slope. The trained Random Forest was then used to classify the remainder of the spikes into the identified single units, noise or non-classifiable units (multi-unit). The selection of auditory units among the spike-sorted ones yielded 1401 spike-sorted auditory units. To further identify single units among spike sorted units, the quality of the spike sorting was assessed both visually by superposing all spike snippets of each unit and quantitatively by calculating a measure of signal to noise ratio (SNR). The SNR measure was defined as the difference between the max and the min of the average spike snippets template, normalized by the standard deviation estimated at those two points across all spike snippets. The SNR values of spike sorted units were compared to the SNR values obtained for a selection of units that could be very clearly identified as single units since their large amplitude and unique shape allowed isolation by threshold. The SNR of these amplitude-selected units were all above 5 and, therefore, we used the SNR to quantify the goodness of our spike sorting and used a threshold of 5 SNR to classify units as single units. Using this approach, we extracted 914 single auditory units from all our recordings.

Decoding the responses of auditory single units.

The goal of this study is to determine whether neural responses reflect the semantic classification of zebra finch calls as performed by Zann (Zann, 1996) and our laboratory, and as inferred both from the acoustical signature of call types and the social context of production. To achieve that goal, neural responses of single units were analyzed using a decoder that used both spike count and spike pattern information (Rieke *et al*., 1995; Borst & Theunissen, 1999; Machens *et al*., 2003; Nelken *et al*., 2005; Schnupp *et al*., 2006; Menardy *et al*., 2012; Menardy *et al*., 2014). Briefly, this decoder found similarities between neural responses to the same or different stimuli by calculating the Euclidean distance between two spike trains of the same length. The spike trains were preprocessed by convolving with a

Gaussian window of varying width, a variant of the van Rossum distance that uses an exponential instead of a Gaussian window (van Rossum, 2001). This process yields for each unit, a confusion matrix of the probabilities to classify spike trains of stimulus i (row) as belonging to the same stimulus or other stimulus j (column). From these probabilities, one can estimate the "information content" of neural responses or more accurately, the Mutual Information (MI) between responses and stimuli. Mutual Information calculations were used to identify units presenting high levels of information about semantics in their responses. The probabilities of classification given by the confusion matrix were also used to estimate the discrimination of semantic categories achieved by an ideal observer of neural responses, investigate the selectivity of this discrimination for one or several semantic categories, and measure the invariance of neural responses within semantic categories.

The decoding analysis was performed on the responses to all monosyllabic vocalizations (Distance, Nest, Long Tonal, Tet and Thuk calls) and on responses to the first sound element of polysyllabic stimuli (Song, Wsst, Distress and Begging calls). The first sound element was defined as any sound longer than 30ms and followed by minimum 10ms of silence before the next sound element. Silences were defined as periods where the absolute value of the waveform was below 35% of the RMS intensity. Neural responses to first vocalization elements were framed into 600 ms analysis windows. This window size ensured that the neural responses (PSTH) of minimum 90.9% of monosyllabic calls and first elements of polysyllabic vocalizations would not be truncated (mean±sd over all single units, 93.4±1.4%; Fig 1). The beginning of the window started at least 10ms before the onset of the first sound element (but could start long before) and ended at least 10 ms after the offset of the sound element. When positioning the window, we also made sure that the window was ending at least 10ms before the onset of the second sound element. If the first sound element was longer than 590ms (9% of the stimuli) then the neural responses were truncated to fit the 600ms window. This cutting process yielded for each unit a list of T neural responses of 600ms length per vocalization (T=10±0.22 is the number of presentations of each vocalization). To estimate if the discharge patterns obtained in response to sounds were all different from the spontaneous neural activity in silence, we also isolated, for each units, 20 x T sections of 600ms neural activity before the beginning of 20 randomly chosen vocalizations. So the total number of Neural Response (NR) sections isolated by this process per unit was NR=(NV + 20)xT with NV the Number of Vocalizations presented to the unit.

For each unit, the spike distance measures were used to compare the NR 600ms-sections of neural responses and decode the presented stimuli. The neural activity was represented as the number of spikes per bin of 1ms, so each spike pattern of 600ms (n=NR) was represented by a binary vector or "word" of 600 elements. Each element in the vector, or "letter" in the "word", indicated the presence of a spike at that particular time point (0 or 1). Then, the algorithm calculated the shortest spike distance between every spike train or "word" and templates of neural responses obtained for each of the NS stimuli (NV vocalizations and 20 silences). The template of neural response of a given stimulus was estimated by averaging all the vectors or "words" but one (T -1), obtained in response to that particular stimulus. The first step to calculate the shortest spike distance between two spike patterns or "words" was to convolve each spike pattern with a Gaussian window of unique width. The width of the Gaussian window was optimized for each unit to obtain the maximum value of Mutual Information in the confusion matrix of categories (see section "Estimating the Discrimination or classification performance of Semantic Informative units"). Nine values were tested (2, 5, 10, 20, 30, 50, 100 and 600ms) and the quartiles of the windows selected for the 914 single units were 30/50/100 ms with an average value of 145.5±185.2ms (mean ± sd). This convolution gave two time-varying mean firing rate responses or "smoothed words" of exact same length. Then every possible delay between the two "smoothed words" was tested to find the shortest Euclidian distance between them. Non-overlapping regions of responses were padded with the time average mean firing rate for each response. This process yielded the shortest distance between each of the NR spike patterns or "words" and each of the NS (NV+20) templates of response to stimuli (vocalizations and silences).

Then, for each neural response, the algorithm predicted the most probable stimulus that had elicited the spike pattern by choosing the stimulus template that had the shortest distance to that spike pattern. The confusion matrix (see an example on Figure 2A) represents for each stimulus i (rows: actual stimuli), the joint probability that the T neural responses to that stimulus i were classified as belonging to the same stimulus (diagonal) or the other stimuli (other columns in that row; columns: predicted stimuli j). A unit that would give robust but different spike patterns to every single stimulus would yield a diagonal matrix of value 1/NS while a unit that would give random spike patterns to every single stimulus would yield a uniform matrix of value $1/NS^2$.

Calculating the Information on semantic categories of single units

From the confusion matrix, one can obtain a measure of the information content of neural responses and estimate the goodness of the classification (effect size) by estimating the Mutual Information (MI) between predicted stimuli and actual stimuli:

$$MI = \sum_{i,j} p(i,j) \times \log_2\left(\frac{p(i,j)}{p(i) \times p(j)}\right)$$

Here the probability of the actual stimulus, p(i), depends on the number of spike trains T obtained for each vocalization while p(i,j) and p(j) are obtained from the confusion matrix and Bayes' theorem respectively. It should be noted that the MI will have a positive bias for small number of events T and is bounded by the number of stimuli tested (NS) (Panzeri & Treves, 1996; Nelken *et al.*, 2005). For example, for NV=34 stimuli (the smallest vocalizations set we used in the present neural recordings), the upper bound for estimates of MI (called from here on MI_{max}) from the confusion matrix is $\log_2(54)$ = 5.75 bits (note that the total number of stimuli NS in the matrix takes into account the 20 silence sections).

The measure of effect size provided by the MI can be used to quantify the amount of information in spike trains and, in this manner, compare results across cells and subjects. However, MI increases with any systematic classification of stimuli and measures any information content of neural responses (e.g. coding for a particular frequency present in the pitch of a group of calls, coding for a duration of calls present in another group of calls, etc.). As such, MI measures cannot be used to estimate any putative neural classifications of groups of vocalizations since the MI does not depend on the positions of the probabilities within a row of the confusion matrix: a matrix for which we shuffle the column positions or labels will have exactly the same value of mutual information as the original matrix. In this study, our aim was to measure the information of neural responses about the semantic categories that we inferred from the behavioral context of vocalization emissions and to compare that value of Categorical Information to the total MI. Furthermore, we distinguished two types of Categorical Informations, the Exclusive Categorical Information (ECI), that measures the information about semantic categories only, and the Inclusive Categorical Information (ICI), that measures both the information about the semantic categories, and the information about each vocalization within the categories. To estimate these two values of Categorical Information, we calculated the mutual information of two modified confusion matrices: the Exclusive-Categorical-uniform matrix (Figure 2C), where we only kept information regarding semantic classification of stimuli and the Inclusive-Categorical-uniform matrix (Figure 2B), where we kept information about both the semantic classification and the classification of single calls within the categories. The Inclusive-Categorical-uniform matrix (Figure 2B) was obtained by keeping intact the probabilities p(i,j) only if i and j belong to the same semantic category and by substituting the probabilities p(i,j) by the average probability of misclassification within each row if i and j did not belong to the same category. Making uniform the distribution of probabilities of misclassification (i.e. classification outside of the same semantic category) was equivalent to maximizing the randomness of predictions outside semantic categories and removed any systematic classification reflecting other types of information content in the neural responses. The Inclusive-Categorical-uniform matrix (Figure 2B) is therefore the matrix with the smallest MI that still preserves all the predictions within categories. In this sense, the MI measure computed this way captures only the information regarding correct semantic classification and correct classification of calls within categories. We call this measure the Inclusive Categorical Information (ICI). Note that the upper bound for estimates of ICI, ICI_{max}, is the same than the upper bound for estimates of MI in the original confusion matrix and equals MI_{max}. For each unit, the value of ICI was normalized by this upper bound for ICI. The Exclusive-Categorical-uniform matrix (Fig 2C) was obtained from the Inclusive-Categorical-uniform matrix by making uniform the probabilities of correct categorical classification (i.e. classification within categories): if i and j belong to the same semantic category then the probability p(i,j) was substituted by the average probability of correct classification within each row. This process maximized the randomness of predictions inside semantic categories and further removed any systematic classification reflecting information about individual calls within categories in the neural responses. The Exclusive-Categorical-uniform matrix (Fig 2C) is therefore the matrix with the smallest MI that only preserves the predictions about semantic categories. We call this measure the Exclusive Categorical Information (ECI). Note that the upper bound for estimates of ECI is smaller than the upper bound for estimates of ICI in the Inclusive-Categorical-uniform matrix or of MI in the original confusion matrix (MI_{max}) and depends on the number of calls per category (NC) as follows:

$$ECI_{\max} = MI_{\max} - \sum_{C=1}^{9} \frac{NC \times \log_2(NC)}{NS}$$

To show our results on a uniform scale, the absolute value of ECI was normalized either by MI_{max} (Fig 4A), by this upper bound, ECI_{max} (Fig4C) or by MI_{Tot}.

Estimating the significance of Exclusive Categorical Information (ECI)

To estimate the significance of ECI for each unit (n=914), we conducted two tests. The first one aimed at finding if the same value of ECI could have been obtained with any categorical classification of the predicted stimuli. This test both controlled for the positive bias due to the small number of events T in the estimation of the ECI and for the various number of stimuli NS used between recording sites. A shuffled ECI was calculated on confusion matrices where all columns except those corresponding to Silence had been shuffled while keeping the semantic labels for the rows and columns intact. For each unit, 1000 shuffled ECI values were obtained (i.e. estimated on 1000 column-shuffled matrices). The actual value of ECI was then statistically compared to the distribution of values obtained by chance for the same unit: if the observed value was found in the upper percentile of the random distribution (corresponding to a p<0.01), the unit was labeled as a Semantic Informative (SI) unit. Note that, in our definition, a SI unit is a unit that provides any information about semantic categories and not necessarily a unit that shows invariance for all stimuli belonging to a category (as in categorical perception). That property will be assessed with our measure of invariance as defined below.

The second test aimed at investigating if the ECI value obtained for SI units (n=404) could be fully explained by the acoustic correlation of stimuli within categories or whether it also depended on non-linear processing of acoustical features that could be captured by the semantic labels of stimuli. To test the relative importance of these linear and non-linear contributions of sound features to the ECI of each SI unit, vocalizations were progressively shuffled between categories, and rows and columns of the confusion matrix reorganized to reflect the new grouping of stimuli. For each level of shuffling, we then calculated a measure of Semantic Disruption that quantified the similarity between the disrupted categories and the true categories. The Semantic Disruption was taken as the proportion of vocalizations not correctly assigned in each new disrupted category. We also calculated a measure of Acoustic Disruption as one minus the average correlation between spectrograms of stimuli in each "new" category. The relationship between Acoustic Disruption and the ECI quantifies the contribution of linear responses to the spectro-temporal features of the sound in explaining the ECI. The additional relationship not explained by the Acoustic Disruption but explained by the Semantic Disruption quantifies the contribution of non-linear responses to the sounds that further contributes to the neural discrimination of call categories. Five hundred levels of intermediate shuffling were tested per SI unit to investigate this relationship. The dependence of ECI on Semantic disruption and/or Acoustic disruption was estimated by comparing the goodness of fit of a quadratic curve to the data with likelihood ratio tests using the

Linearmodel.fit function of Matlab. To estimate the relative importance of the two parameters (Acoustic disruption vs Semantic disruption), the full model constructed with the two parameters was compared to the models taking into account only the Acoustic disruption or the Semantic disruption. Based on the significance of the comparison between the full model and the model taking into account only the Acoustic disruption (significance threshold p=0.01), we distinguished 2 different types of SI units: when the likelihood ratio test was significant, then SI units were labeled as “AS>A” units indicating that the acoustic predictor did not provide information that was not explained by the semantic predictor and the semantic predictor provided additional information; when the full model was not different from the model constructed with the acoustic predictor, then SI units were labeled as “AS=A” units, indicating that semantic content did not provide information that was not explained by the acoustic predictor, with the implicit understanding that acoustic prediction means here predictions based on linear responses to spectro-temporal features.

Estimating the Discrimination or classification performance of Semantic Informative (SI) units

In addition to using ECI as a measure of semantic discrimination or classification, we also quantified discrimination for each SI unit by directly using the probabilities of the confusion matrix. For each SI unit, the discrimination of a given semantic category was estimated by the percentage of correct classification (PCC) of the vocalizations belonging to that category. The PCC value of category C was calculated by first adding up the joint probabilities of any vocalization belonging to C and predicted as belonging to C, in other words, by first adding up the joint probabilities within the block diagonal corresponding to C in the confusion matrix. That sum was then converted to conditional probability by dividing by the sum of the number of neural responses (T) to all the vocalizations of C used to construct the confusion matrix. This process yielded for each unit one value of PCC per semantic category and an average discrimination performance of semantic categories by the unit: the mean of these PCC values over semantic categories. To evaluate if the classification of vocalizations within a category by a given SI unit was significantly higher than chance, a binomial test was run using the binocdf function of Matlab. If the p-value of that test was below 0.01 then the SI unit was considered as Discriminant for that semantic category. All SI units but one significantly discriminated at least one semantic category.

Estimating the Selectivity of SI units

SI units could discriminate all categories evenly or, at the other extreme, selectively classify only a single category against all the others. To investigate such different levels of selectivity, we used two selectivity measures based on the PCC values: the Global Selectivity (GS), to quantify overall or average selectivity and the Selectivity measure (Sel) calculated for each semantic category. The GS was based on an entropy measure of PCC across all categories. The entropy measures how close the distribution of PCC values is from a uniform distribution: a non-selective unit would have the same value of PCC for all 9 categories and a high value of entropy of PCC while a highly selective unit would have a high value of PCC for one category and low values for the others, yielding a low value of entropy. GS was defined as follows:

$$GS = 1 - \frac{H_{obs}}{H_{\max}}$$

with H_{obs} the observed entropy calculated on PCC of categories and $H_{\max}$ the maximum possible entropy obtained if all PCC values had been equal. The entropies were calculated as follows:

$$H_{obs} = \sum_{c=1}^{9} -PCC_c \times \log_2\left(PCC_c\right)$$

$$H_{\max} = \sum_{c=1}^{9} -\frac{1}{9} \times \log_2\left(\frac{1}{9}\right) = -\log_2\left(\frac{1}{9}\right)$$

Note that PCC values were normalized so that their sum would be 1 before calculation of entropy values.

Sel measured the selectivity of the classification performance for each semantic category and was defined as the following ratio:

$$Sel_{c=i} = \log_2\left(\frac{8 \times PCC_{c=i}}{\sum_{\substack{c=1 \\ c \neq i}}^{9} PCC_c}\right)$$

Sel measured the factor by which a given semantic category was better (Sel>0) or not as well (Sel<0) classified compared to the average of other categories. A unit was considered highly selective for a given category if the Sel value for that category was higher than 1.75. This threshold of 1.75 was chosen as the value above which only 5% of the SI units would present two semantic categories with values of Sel above the threshold; in other words, the probability for a unit to be selective for two semantic categories was below 0.05 (SuppFig3).

Estimating the Invariance of neural responses of SI units

For any SI unit, invariance of neural responses was estimated for each category significantly discriminated by that unit, by the Invariance index (Inv). Inv is based on the entropy of the joint probabilities of the confusion matrix and is calculated as follows:

$$Inv_C = \frac{H_{C,obs} - H_{C,\min}}{H_{C,\max} - H_{C,\min}}$$

with $H_{C,obs}$ the observed entropy of the joint probabilities p(i,j) for i and j belonging to the semantic category C:

$$H_{C,obs} = \sum_{i,j \in C} -p(i,j) \times \log_2(p(i,j))$$

with $H_{C,\max}$ the maximum possible entropy of the joint probabilities p(i,j), for i and j belonging to semantic category C and given that V vocalizations belong to that category:

$$H_{C,\max} = -\log_2\left(\frac{1}{V^2}\right)$$

and with $H_{C,\min}$ the minimum possible entropy occurring when every vocalization is exactly correctly classified within the category although some misclassification outside of the category are allowed:

$$H_{C,\min} = \sum_{i \in C} -p_j(i) \times \log_2\left(p_j(i)\right) \qquad p_j(i) = \sum_{j \in C} p(i,j)$$

Note that p(i,j) values of each category C were normalized so that their sum would be 1 before calculation of entropy values.

The rationale was that a unit highly invariant for a given category C would respond with the same neural response to all vocalizations of that category and would thus have a uniform distribution of probabilities p(i,j) in the confusion matrix for i and j belonging to C, and as such, a high entropy value. Conversely, the least invariant unit for a given category C would respond with distinct neural responses to all vocalizations in that category and would thus have a non-uniform distribution of probabilities p(i,j) and therefore a low value of entropy, in the confusion matrix for i and j belonging to C. In other words, highly invariant SI unit would display a uniform distribution of probabilities in the block of the confusion matrix corresponding to that category while least invariant SI units would display a diagonal in that same block.

The global invariance of SI units was estimated by calculating the mean Inv value over the categories for which the SI unit was significantly discriminant.

Statistical tests

Differences of unit proportions were analyzed with Chi-square tests. Correlations between PCC, Inv, Sel and ECI were estimated by calculating the Spearman rho with the *corr* function of Matlab. For each SI unit, the correlation between its performance on the classification of the vocalizations in the different semantic categories and those of the PCLDAS was estimated by calculating the Spearman rho with the *corr* function of Matlab. The significance of the effect of semantic categories or auditory regions on values of PCC, Sel and Inv was evaluated by tests of Kruskal-Wallis (KW). Note that for the effect of regions, only the 6 auditory regions L1, L2, L3, CMM, CLM and NCM were taken into account for the statistical tests although results of units of unknown positions are also shown in figures. Comparisons between primary auditory regions (L1, L2 and L3) and secondary auditory regions (CMM, CLM and NCM) were performed using Mann-Whitney-Wilcoxon

(MWW) tests with the *ranksum* function of Matlab. Note that CLM is sometimes grouped with the primary auditory area, which is then called the Field L/CLM complex (Wang *et al*., 2010)**.** Finally the PCC of the PCLDAS and the PCC of Discriminant SI units for the different semantic categories were compared with an N-way analysis of variance (ANOVA, *anovan* function of Matlab) followed by Tukey-Kramer post-hoc tests. All computations and statistical analyses were done under Matlab.

RESULTS

The goal of our study is to understand how the information about semantic categories is encoded in single neurons in the avian auditory cortex. To quantify and identify the nature of this semantic information, we used a decoding approach based on similarities between spike patterns obtained in response to a large database of vocalizations. This vocalization bank included almost all of the vocalizations of the zebra finch repertoire (see methods). To investigate the neural representation of semantic categories using this decoding approach, we identified whether groups of vocalizations eliciting similar spike trains would reflect the semantic categories expected from the ethological relevance of the vocalizations. This led us to a definition of Exclusive and Inclusive Categorical Information (ECI, ICI) that we used to identify neurons that code semantic categories. We also used the Exclusive Categorical Information as well as the probabilities of correct classification of vocalizations in semantic categories to quantify neural discriminability for semantic information. To further explore how semantic information is encoded, we first examined the extent to which semantic categorization could be expected from linear responses to spectro-temporal auditory features. Then we examined the nature of this neural categorization by quantifying neural selectivity and neural invariance. Neural selectivity is used to examine the degree with which single neurons discriminate one category over all others. Neural invariance quantifies the similarity between neural responses to different vocalizations in a given category.

Example Responses

Coding for semantic categories was reflected in the response of single units both in primary (Field L) and secondary auditory areas (NCM, CMM, CLM; note that CLM is sometimes grouped with the primary auditory area which is then called the Field L/CLM complex, (Wang *et al*., 2010)) but neurons varied strikingly in their degree of selectivity and invariance. In Figure 1, we show representative responses of 4 single units chosen to illustrate some of this variability. All of these example units were classified as Semantic Informative units (or SI units; see methods and below) and had PCC greater than 0.4 for at least one vocalization type. The units shown in panel A and B are examples of highly selective units, here for Wsst calls and Distance calls, respectively. In both cases, the firing rate is much higher for the two examples of Wsst and Distance calls shown and the same increase in responses was observed for all calls in those category (not shown on the figure). Note also that beyond having higher firing rates, these units respond with reliable spike patterns to these stimuli suggesting that both rate and spike patterns can play a role in the encoding of semantic categories. The example unit in panel B (selective for Distance calls) also shows responses, albeit weaker, to Long Tonal calls and to one Song syllable. These responses that are “off” the preferred category can be explained in this case: Long Tonal calls are the precursor of Distance calls, produced by juveniles and songs often contain elements matched to the bird’s Distance call.

The single unit shown on panel C has an intermediate level of selectivity as it responds with increased firing rates to Wsst and Distance calls and to a lesser extent to Thuk, Distress Calls, Long Tonal Calls and Song syllables. Such a unit can clearly code information about the semantic categories but in a more distributed coding scheme. Finally, the unit on panel D responds to vocalizations in all categories. Although its average firing rate might not be very informative for the decoding of semantic categories, it clearly exhibits reliable spike patterns to specific sound features and such responses can be used to successfully decode sound identity and perform some degree of categorization. Such a neuron could then participate in semantic information processing but using a distributed coding scheme based on spike patterns.

In the rest of the results, we will quantify such coding properties across the population of single units using our measures of category discrimination, selectivity and invariance. We will also describe how these coding properties vary across regions of the avian auditory cortex.

Information about Semantic Categories

The decoding of stimulus identity based on spike train comparison yields for each single unit a confusion matrix of probabilities of stimulus classification (see examples of confusion matrices in Fig 2A and Fig 8). Each row i of the confusion matrix gives, in the different columns j, the probabilities of classifying the neural responses to stimulus i as obtained in response to the same stimulus (j=i) or to the other stimuli (j≠i). From the confusion matrix, we derived a metric quantifying how much information the unit is encoding solely about the semantic category of the vocalizations, the Exclusive Categorical Information or ECI (see methods and Fig 2C). Across all auditory units (n=914), we obtained an average ECI of 0.20±0.005 bits (mean ± standard error). To test whether the ECI of each unit could have been obtained by any categorization of vocalizations, i.e. a categorization other than semantic, this ECI value was compared to the distribution of bootstrapped-ECI values obtained by shuffling the semantic labels of the vocalizations in the columns of the confusion matrix (see methods and Fig 3A). Out of the 914 single units recorded in the auditory cortex of 6 anesthetized birds, 404 have values of ECI higher than what could be expected by chance (Fig 3B). Such units, called Semantic Informative (SI) units throughout the present study, have an average value of ECI of 0.28±0.008 bits (mean ± standard error; max value 1.59) while Semantic Uninformative units' average value is 0.13±0.003 bits (mean ± se).

Values of information are constrained by the size of the stimulus set. To estimate the proportion of information that was dedicated to semantic categories for each unit, values of Total Mutual Information (MI) and ECI and ICI were normalized for each unit by the maximum value of MI that could be expected given the dataset (see methods). These normalized information values are shown on Fig 4 where ECI and ICI are plotted against MI (Fig. 4A and 4C, respectively). For validation and comparison, we also show on the same plots the values of ECI and ICI that could be expected from neurons with exactly the same information about stimulus identification (same MI) but random groupings of these stimuli (Average Chance Value points on the plots). Clearly, many neurons in the avian auditory system have values of ECI that are greater than those obtained by chance: the SI units. SI units had an average value of MI of 37.1% ± 0.5% (mean ± se) of the maximum possible value. On average, 34.3% ± 0.6% (mean ± se) of that information was about both semantic categories and individual vocalization within categories as measured by the ICI, while 11.8% ± 0.4% (mean ± se) was solely about semantic categories as measured by the ECI. Although that fraction of information dedicated to coding semantic category might appear relatively small, one should remember that the ECI is also limited by the number of semantic categories and the composition of the dataset (see methods). To further quantify in absolute terms the amount of information for categories relative to the absolute maximum, we also show the data after normalizing the ECI by the maximum achievable value of ECI given the dataset (Fig 4B). The average value of this normalized ECI is 9.1% ± 0.3% (mean ± se) for SI units and although many SI units code above 10% of the potential maximum, it is clear that a full categorization of semantic classes will require the decoding of an ensemble of neurons.

The SI units were found in all of the 6 auditory areas investigated in this study (three Field L regions: L1, L2 and L3; two Mesopallium Caudale regions: Mesopallium caudolateral, CLM, and Mesopallium caudomediale, CMM; and Nidopallium caudomediale, NCM; Fig 5A & 5B). However, the proportion of SI units is significantly different between regions (X^2_5=17.5, p<0.01, Fig 5D) with Field L regions and CMM presenting higher proportions of SI units than NCM and CLM.

Besides the measure of information about semantic categories, the discrimination performance can also be quantified with the percentage of correct classification (PCC) of vocalizations within each category (see methods). The average PCC value is calculated across all categories and is highly correlated with the ECI (Spearman rho=0.71, p<0.001; SuppFig 4A & 4D). The PCC of each semantic category was used to determine category dependent discrimination performance. All SI units, except one, correctly classified above chance level (p<0.01) the vocalizations of at least one semantic category (mean ± sd numbers of categories with PCC above chance: 4.23±2.14; range: 1-9) with PCC values ranging from 14.3% to 60% (5% and 95% quantiles) and as high as 92% (see Fig 7A).

Contribution of Linear Auditory coding for coding of semantic categories

Vocalizations belonging to the same category share a large amount of acoustic properties (Fig Supp2), and a linear discriminant analysis applied on spectrograms (PCLDAS, see methods) shows that very good classifications can be obtained based on a linear combination of spectro-temporal features (see Fig 7B). All of our SI units are auditory and produce distinctive spike patterns in response to specific auditory features. The information provided by SI units for semantic categorization is either based on linear or non-linear neural tuning to such acoustical features. Although the goal in this study was not to estimate such non-linear functions, we wanted to estimate the degree to which linear responses to specific spectro-temporal features could contribute to semantic information. Indeed, the linear tuning for these features could be exactly redundant with a semantic classification (indicating that the linear response was perfectly tuned for the discrimination of semantic categories). Alternatively, adding semantic codes to the linear auditory tuning could yield additional information (indicating that a fraction of the non-linear response to sounds was also tuned for discrimination of semantic categories).

For this analysis, we used a shuffling procedure that quantified the sensitivity of the ECI both to the linear correlation between acoustic features of calls belonging to the same category, and to the integrity of semantic labeling of calls within the same category (see the methods section “Estimating the significance of Exclusive Categorical Information”). Calls were progressively shuffled between categories so as to create chimeric categories with lower values of acoustic correlation between calls (higher levels of Acoustic disruption, Fig 6A) and lower proportions of calls with the same semantic labeling (higher levels of Semantic disruption, Fig 6A). A new ECI was then calculated for each of these row and column-reorganized confusion matrices. As exemplified in figure 6A, both Acoustic and Semantic disruptions are good predictors of ECI with adjusted R-squares of 0.57±0.19 (mean ± sd) for Acoustic disruption and 0.60±0.20 (mean ± sd) for Semantic disruption. To compare the relative importance of Acoustic and Semantic disruptions as predictors in the model, we estimated the increase in goodness-of-fit when combining both predictors in a third model.

For 14.1% of SI units, adding the Semantic disruption as a predictor to the Acoustic disruption in the model does not significantly increase its goodness of fit. These "AS=A" units (Fig 6B & 6C) are performing a classification of calls in semantic categories that can be explained to its full extent by their linear response properties to sounds. However, for 85.9% of SI units, the Semantic disruption predictor does provide additional information that is not explained by the Acoustic disruption predictor and significantly increases the goodness-of-fit of the model from values of adjusted R-squares of 0.61±0.01 (mean ± se) to 0.64±0.01 ("A+S>A" units in Fig 6B & 6C). In other words, the categorization of calls made by most of SI units cannot be explained to its full extent by the linear responses to spectro-temporal acoustic properties of the vocalizations. By comparison, for 84.9% of SI units, adding the Acoustic disruption as a predictor to the Semantic disruption in the model does *not* significantly increase its goodness-of-fit ("A+S=S" units in Fig 6C). Thus our semantic disruption coefficient clearly captures efficiently both the linear and non-linear response contributions to the ECI.

The two populations of SI units that can be distinguished based on the dependence of their Exclusive Categorical Information to Semantic disruption (the non-linear "A+S>A" and the linear "A+S=A") are evenly distributed in the 6 auditory regions investigated here (X^2_5=10.1, p=0.07, Fig5 C&D). So in all regions the large majority of SI units are performing a classification of vocalizations into semantic categories that cannot be explained solely by linear responses to spectro-temporal acoustical features of specific categories. We will exemplify these properties in the following section when we examine the firing rate of selective SI neurons as a function of spectro-temporal tuning along the dimensions that maximally separate one class from all others.

Selectivity

SI units significantly discriminate at least one category above chance but can also discriminate multiple categories. To investigate degrees of selectivity and distinguish highly selective SI units (discriminative for one particular semantic category) from non-selective units (discriminative for more than one semantic categories), we quantified selectivity based on the performance of classification of each category. Our measure of selectivity, Sel, is unit and category dependent: it compares the PCC value obtained for a particular semantic category to the average PCC value obtained for all other categories. Positive values indicate that the category is better classified than the others while negative values indicate that it's not as well classified. For each unit, Sel was only calculated for the categories discriminated above chance. SI units show various degree of selectivity for the category(ies) that they discriminate, from non-selective (Sel<0 or x1 on Fig 7A) up to highly selective (Sel>1.75 or x3.36 on Fig 7A) with most values of Sel between -0.6 and 2.2 (5% and 95% quantiles). To illustrate that high values of Sel correspond to robust and distinct neural responses to a particular call type, we show on figure 8 two examples of non-linear neural responses of highly selective units (Sel>1.75). The x-y plots on the right panel show the firing rate of these units versus the stimulus described by its spectrogram and projected onto the spectral-temporal dimension that best distinguish that category from all other semantic categories. These acoustic dimensions are obtained using logistic regression on the spectrograms from all the vocalizations in our database (see methods). Although this representation only shows a neural code based on spike counts (mean rate) while the classification performances of the units are estimated using both spike counts and spike patterns, one can see that these units show categorical responses to semantic calls: units switch non-linearly from low spike-rates to high spike-rates as soon as the stimuli pass a threshold in the linear acoustic dimension that best discriminates the semantic category. For all semantic categories, Sel values were positively correlated with PCC values (see Table 2, Fig 7A), indicating that performance of classification of a given category is increasing with units' selectivity for that same category. The selectivity of SI units was also estimated using the Global Selectivity index, a measure based on the entropy of the percentages of correct classification for all categories (see methods). The relationships between that measure and the information measures MI, ICI and ECI can be found in SuppFig 4B & 4E.

Invariance

Finally, we wanted to quantify the degree with which neural responses were invariant or tolerant within the categories significantly discriminated. Indeed, a unit that is significantly discriminating the vocalizations belonging to a semantic category (for example by showing a robust and distinctive firing rate to all calls in that category compared to calls from another category) could exhibit different responses for each call within that category (such as different spike patterns). Such neurons would have a low level of invariance. On the other hand, a highly invariant unit would show the same spike count and pattern in response to all the different calls of the same semantic category. High invariance could therefore reflect specialization for semantic categorization.

The invariance of units for a given semantic category was quantified using the Invariance Index (Inv), which measures how uniform the probabilities of classification of calls are within the same category (see methods). An Inv value of 1 corresponds to a unit that gives very similar spike trains in response to all the calls of the category, such a unit would show a uniform distribution of probabilities in the block of the confusion matrix corresponding to that semantic category (see Fig 8A for a unit highly selective and highly invariant for Wsst calls). An Inv value of 0 indicates a unit that gives distinguishable spike trains for each call within the category and such a unit would produce a perfect diagonal in the block of the confusion matrix corresponding to that category (see Fig 8B for a unit highly selective and poorly invariant for Distance calls). Most of the SI units (95%) showed values of Inv between 0.05 and 0.47 for the categories they were significantly discriminating and could therefore be classified has having low-invariance rather than high invariance (see Fig 7A). However, some units had values of Inv up to 0.96. For all semantic categories except Distress calls, Inv values are positively correlated with PCC values (see Table 2, Fig 7B), revealing that performance of classification of a given category is increasing with units' invariance for that same category. Finally, the global invariance of SI units was estimated by averaging Inv over categories for which they were discriminant. The relationships between that measure and the information measures MI, ICI and ECI can be found in SuppFig 4C & 4F.

Coding Across Semantic Categories

Although we found examples of discriminating, selective and invariant units in all categories, the distribution and the mean of these coding measures varied drastically across semantic categories. First, although for each semantic category, one can find a subset of SI units that correctly classify the vocalizations above chance (see methods), the number of these Discriminating SI units is higher for Wsst calls, Distance calls and Song syllables (X^2_{17}=263.3, p<0.001) than for other call categories (see inserts in Fig 7A). The average performance of discrimination, the average selectivity and the average invariance of these discriminating SI units are also higher for the three latter call types (Fig7 B-D; PCC: Kruskal-Wallis X^2_8=451.9, p<0.001; Sel: Kruskal-Wallis X^2_8=532.6, p<0.001; Inv: Kruskal-Wallis X^2_8=320.5, p<0.001). To test if differences in discrimination between semantic categories are due to some intrinsic bias in the discriminability of calls, the classification performances of units are compared to those obtained from Linear Discriminant Analysis applied on the spectrogram, the PCLDAS (see methods). This classifier finds the spectral-temporal parameters of vocalizations that can be used to optimally discriminate all semantic categories.

For every category, except Distress calls, the PCLDAS shows higher classification performance than the average PCC obtained for the Discriminant SI units (Fig 7B, Anovan, F_1=1272.3, p<0.001, Tukey-Kramer post-hoc test all p<0.001 except Distress calls). Note however that particular units can have PCC values close to or even above the values obtained in the PCLDAS (Fig 7A). Higher values are possible since the PCLDAS is a linear classifier based on spectro-temporal features and neurons could act as non-linear classifiers. While the performance of the PCLDAS is different between call types (KW, X_8=75.3, p<0.001), the semantic categories best classified by the PCLDAS are different from those best classified by the SI units (Fig 7B): for any given SI unit, there is no correlation between the performances of the PCLDAS and the performances of classification of the various call types (mean ±se Spearman rho: 0.04±0.02; all p>0.01). So the higher classification performance of Wsst call, Distance call and Song by SI units cannot be attributed to easier acoustic discriminability.

Anatomical Specialization

We studied whether discrimination, selectivity and invariance properties of SI units were different between avian cortical regions. The average classification performance of units, measured by the mean PCC over categories, is higher in primary auditory regions (Field L sub-regions: L1, L2 and L3) compared to secondary auditory regions (CMM, CLM and NCM; Effect of the 6 regions: KW, X_5=26.0, p<0.001; comparison primary *vs* secondary regions: MWW, RS=12849, p<0.001; see Fig 9A & 9G). Within some of the semantic categories, the performance of classification is also different between regions with Wsst, Distress calls and Song Syllables being better classified in Field L regions than secondary auditory regions (Fig 9D, Ws 6-regions KW, X_5=32.8, p<0.001; Ws primary vs secondary MWW, RS=12667, p<0.001; Di KW, X_5=18.2, p<0.01; Di MWW, RS=12672, p<0.001; So KW, X_5=30.5, p<0.001; So MWW, RS=13278, p<0.001), while Distance calls are better classified in L3 and CMM compared to other regions (Fig 9D, KW, X_5=15.1, p<0.01; see Fig 8B for an example CMM unit). The selectivity of each unit is estimated here using both the measure Sel estimated for each semantic category and a second global measure (GS, Global Selectivity, see methods) that is based on the entropy of the percentages of correct classification for all categories. The SI units in the six auditory regions had similar average values of GS (Fig 9B; KW, X_5=3.5, p=0.63) indicating that none of the 6 auditory regions studied here could be considered as more selective. However, investigating the Sel values of Discriminating SI units per semantic category revealed that units that significantly discriminated Distance calls were the most selective in CMM (Fig 9E&9H; KW, X_5=17.0, p<0.01) while units that significantly discriminated Nest calls tended to be the most selective in CML and NCM (Fig 9E&9H; KW, X_5=11.7, p=0.04). Similarly, an average invariance value did not show anatomical specialization but invariance values for specific calls did. Indeed, the invariance values of SI units averaged over categories for which they were discriminant were similar between the different regions (Fig 9C&9I, KW, X_5=10.4, p=0.07). However, SI units discriminant for Nest calls or for Wsst calls tended to be the most invariant in L1 (Fig 9F; Ne KW, X_5=13, p=0.02; Ws KW, X_5=11, p=0.05; see Fig 8A for an L1 unit selective and invariant for Wsst calls).

In summary, Wsst calls, Distance calls and Song Syllables are particularly well discriminated and Distance calls are best classified in L3 and CMM while Wsst calls and Song Syllables are best classified in the primary auditory regions. Neurons in CMM also show higher selectivity for Distance Calls, and higher invariance for Wsst calls and Nest calls is found in L1.

DISCUSSION:

In this work, we performed the first comprehensive investigation of how the meaning of vocalizations used for communication is represented in the avian auditory cortex. Such investigation required a very large data set of stimuli in order to include many examples of the majority of vocalizations types produced by zebra finches, emitted by various male and female vocalizers at various ages. Although responses to tets, distance calls and songs have been examined in previous work (Theunissen *et al*., 2000; Grace *et al*., 2003; Amin *et al*., 2007; Beckers & Gahr, 2010; 2012; Menardy *et al*., 2012; Menardy *et al*., 2014), this is the first time that neural responses to the majority of calls found in the vocal repertoire of this species has been studied.

To examine the neural representation of "meaning", we paid particular attention in the design of our study to include sufficient examples of vocalizations from a given category, both from the same individual (renditions) and from different individuals (or vocalizers). By ensuring that our stimulus set included sufficient variability, we were able to distinguish coding for the semantic category of the vocalization from coding for other information present in the calls. Clearly limitations in neural recording time and other practical considerations prevent the investigation of all naturally occurring variability (e.g. propagation, noise, multiple voices). However, given that the upper bound for mutual information obtained from our single unit responses to our stimuli was not reached, we are confident that we investigated the coding properties of neurons for semantic categorization with a representative and sufficiently large dataset of stimuli. We analyzed the neural representation of these communication calls using a decoding methods based both on spike count and spike patterns. As revealed both by direct measure of the semantic information content of neural responses using information theoretic measures and by the performance of category discrimination of the decoder, the neural responses recorded in all the regions of the avian auditory cortex contain Exclusive Categorical Information. For most of the units, this Exclusive Categorical Information cannot be solely explained by linear tuning to spectro-temporal features (Fig 6). The non-linearities are clearly exemplified in the highly selective units that can show step-wise changes in their response rates to sound even in the linear acoustic dimension that best isolates the category they are selective to from other categories (Fig 8). The positive correlations between discrimination performance, selectivity and invariance indicate that for any semantic category, better discrimination is achieved by an increased selectivity for that category and an increased invariance to the acoustic variations between renditions or vocalizers of the same vocalization type. We also found that the 9 semantic categories investigated in the present study are not evenly represented in the auditory cortex of zebra finches: Wsst calls, Distance calls and Songs are discriminated by more units and with higher values of discrimination, selectivity and invariance compared to other semantic categories. An acoustical analysis shows that this difference in the neural discrimination of call categories cannot be explained by some categories having more idiosyncratic spectro-temporal features than others. Finally, we found differences in coding for semantic categories across avian auditory cortical regions. Field L regions (Primary auditory cortex) seem to encode more information about the semantic classification of the communication sounds than secondary regions; Field L contains higher proportion of Semantic Informative (SI) units and these units have higher average values of discrimination of semantic categories. The caudomediale Mesopallium (CMM) might have a particular role in the representation of Distance calls since these calls are very well discriminated in that region with highly selective units.

Distributed or sparse code for meaning

To achieve the perception of “meaning”, an animal should first be able to discriminate between the spectro-temporal characteristics of vocalizations with different meaning and show differential neural responses to each vocalization category. This neural discrimination could be, on the one hand, achieved by idiosyncratic responses of single neurons to each vocalization type (a distributed code) and/or, on the other hand, by the robust activation of distinct subset of neurons in response to each vocalization category (a sparse code). Neurons that show robust responses for only one category of vocalizations and random spiking behaviors for other categories would be qualified as selective for that particular category. In our study, semantic discrimination is revealed at the level of single neurons both by the Exclusive Categorical Information metric and by the percentage of correct classification achieved by an ideal observer of neural responses. These discrimination measures attempt to identify neurons showing idiosyncratic responses to each category. In previous research, discrimination properties of single units have often been investigated using d-primes calculated on spike rates (Theunissen & Doupe, 1998; Grace *et al.*, 2003; Amin *et al.*, 2007; Menardy *et al.*, 2012; Menardy *et al.*, 2014). While this measure can reveal neural discrimination between two categories based on spike rate, it has also often been used as a measure of selectivity. In our study, however, selectivity is reserved to identify sparse coding, i.e. units that would only encode information about one or a small number of categories. Our results show that the representation of communication calls is mostly distributed but that a continuum is found between units using a distributed code and units using a sparse code. In other words, contribution to semantic information is achieved both by highly discriminating units with low selectivity and highly selective units that show high levels of discrimination for only one or two categories. The distribution of these different neural representations is not uniform through the auditory cortex: contrary to secondary auditory regions, primary auditory cortex units achieve better average classification over all categories, i.e. code more information about all semantic categories. Similar findings have been described in Starlings where two Field L regions (L1 and L2) were shown to exhibit low selectivity and to be responsive to various song motifs while downstream regions (CMM, NCM, but also L3) were more selective for particular song motifs (Meliza & Margoliash, 2012). Similarly in the Macaque, the Superior Temporal Gyrus (STG, a secondary auditory region) codes more information about conspecific vocalizations than the vlPFC to which it projects (Russ *et al.*, 2008). All these results are consistent with the hypothesis that neural coding becomes sparser as one ascends the auditory pathway.

Invariance of neural responses in the bird auditory cortex

To achieve categorization of meaningful stimuli, neural responses should not only be different between categories but also invariant to the variations of the spectro-temporal features that encode other information such as the identity of the vocalizer (Tsunada &

Cohen, 2014). In our study, we were able to measure the degree to which units discriminating a particular semantic category could be invariant to the variations of production of that particular vocalization type. Contrary to most previous studies, the acoustic variations within categories were due to both the variations one can find between renditions of the same vocalization by the same individual and the variations of production between different vocalizers emitting the same vocalization (Simpson & Vicario, 1990; Vicario *et al.*, 2001; Vignal *et al.*, 2004b; Perez *et al.*, 2012). Moreover, instead of investigating the invariance properties for a single vocalization category (Meliza & Margoliash, 2012), we measured invariance of unit responses across all categories that were significantly discriminated above chance. On the one hand, we found that both in primary and secondary auditory areas, one could find some SI units that exhibited very high values of invariance. These invariant properties could reflect the output of a series of auditory computations performed by ensemble of neurons that would be lower in the auditory processing stream. On the other hand, the large majority of the single units in these avian auditory cortical areas had moderately low values of invariance. Thus, semantic categorization as reflected by a high values of invariance might begin at the level of the avian auditory cortex but might also occur in other brain regions analogous to the prefrontal cortex of Mammals (Tsunada & Cohen, 2014). Indeed, in Macaques, several studies have shown the representation of abstract categories such as the quality of the food in food-related calls (Gifford *et al.*, 2005a; Cohen *et al.*, 2006) or the number of auditory stimuli (Nieder, 2012) to happen in the vlPFC. Future recordings of neural responses to conspecific vocalizations in the avian prefrontal cortex, the Nidopallium caudolaterale (Gunturkun, 2005; 2012) might reveal how categorical representation of meaning is achieved in the brain of birds. Finally, a recent study suggest that the vocalizer-invariance in the response of NCM neurons to distance calls is influenced by anesthesia (Menardy *et al.*, 2014). Since our recordings were performed under urethane anesthesia, we cannot rule out the possibility that discrimination and invariance properties of neurons might be different in awake animals.

Disentangling semantic categorization from acoustic categorization

Romanski and Averbeck suggested that one way to demonstrate semantic categorization in the brain would be to show that sounds with similar acoustic morphology but different semantic context evoke different neural responses (Romanski & Averbeck, 2009). Interestingly, the activity of neurons in the male mouse basolateral amygdala in response to the same female vocalization is modulated by the context (predator cue vs mating cue) while it's not when the sound presented in these two contexts is a burst of noise (Grimsley *et al*., 2013). This differential neuronal activity is correlated with the male mouse behavior in response to the same vocalization (escape in presence of the predator cue and approach in presence of the mating cue). Similar modulation of neuronal activity in response to the same conspecific vocalizations has also been demonstrated in the secondary auditory cortex of the zebra finch: while the firing rate of NCM neurons can be used to discriminate between familiar and unfamiliar distance calls when the bird is accompanied by peers, the spike rate is undistinguishable between the two same stimuli when the bird is isolated (Menardy *et al*., 2014). While these results might be explained in terms of semantic categorization or representation, they might also reflect modulation of auditory responses by other sensory modalities. For example, the meaning of a mouse call (mouse in pain) might be unchanged in the two contexts while the hedonic value of the stimulus evaluated by the mouse in regard to other sensory modalities output is changing and so the evoked activity in the amygdala is also modulated. In the case of the results observed in the zebra finch, social isolation is a stressful trigger that could modulate the responsiveness of NCM to conspecific calls of any kind.

Another way to disentangle semantic categorization from acoustic categorization is to study if neural responses are linearly related to the presence of particular acoustical features in the vocalizations, in other words, if neurons act as simple acoustic filters. This is the approach we used here. We investigated to which extent the capacity of neurons to provide information on semantic categories was solely dependent on the acoustic correlation between vocalization or also depended on non-linear acoustic responses that could be captured by the semantic label of vocalizations. We found that the coding of semantic categories of most SI cells in the primary and secondary avian auditory cortex cannot be explained solely by acoustic similarities between vocalizations. Instead responses to calls from different classes are better explained by a semantic labeling showing that non-linear responses are in part tuned for semantic categories. To further study the nature of these computations, future studies should directly estimate how much semantic categorization of vocalizations can be obtained in forward encoding models based on linear spectro-temporal receptive fields (STRFs) and their non-linear enhancements (Theunissen & Elie, 2014). In particular, one could test whether a non-linear encoding model that includes vocalization category enhances the prediction of simpler STRF models.

In the present study, we also based our semantic labels on behavioral and acoustical observations performed by us and others both in the lab and in the field (Zann, 1996; Elie *et al*., 2010; Elie & Theunissen, in prep). Birds might actually categorize calls differently than these human experts either by making different boundaries across categories (finer or coarser) or by a hierarchical classification system (e.g. affiliative *vs* non-affiliative at the higher level, aggressive *vs* distress at a lower level). These alternative schemes for classification would also be reflected in neural responses and future decoding analysis using unsupervised classification approaches could be used to test such hypotheses.

Bias in the representation of Distance calls, Wsst calls and song

Interestingly, the nine vocalization categories studied here were not evenly represented in the neural responses of SI neurons: Distance calls, Wsst calls and Songs were discriminated by more neurons than calls in the six other categories. Higher values of discrimination, selectivity and invariance were also obtained for these three categories. This bias in neural representation could not be explained by differences of discriminability for these categories based on spectro-temporal acoustical features: the average performance of discrimination of the PCLDAS for the different categories was not correlated with the average performance of neural classification by SI units. The bias in the tuning of auditory cortex for these three categories could be indicative of biological constrains related to the particular significance of these vocalizations. Indeed, songs and distance calls encode the identity of the vocalizer (Zann, 1996; Vicario *et al.*, 2001; Vignal *et al.*, 2004b) and play a major role in the reproductive and pair-bonding behavior of the zebra finch. Both vocalizations are used for individual recognition and recent studies have implicated secondary auditory areas, in particular NCM, in the discrimination of familiar vs stranger vocalizers based on the Distance calls (Zann, 1996; Vicario *et al.*, 2001; Vignal *et al.*, 2004b). Songs are learned from a tutor and emitted by males to attract females and synchronize copulation (Zann, 1996). Distance calls are essential contact calls that adults emit to re-establish acoustic contacts with their partner, offspring or other familiar individuals when those are out of sight (Zann, 1996). Distance calls also convey some emotion-like information about the vocalizer (Perez *et al.*, 2012). In a nutshell, Distance calls and songs are signals that convey a lot of information and the extraction of this information could require additional computations compared to other call categories. However, the potential of other call categories, including the Wsst call, to encode the identity of the vocalizer or other type of information has never been investigated, neither by acoustic analysis of the calls nor in behavioral experiments. Thus, the neural specialization observed here for Distance call, Wsst call and songs needs further explorations.

The representation of the Distance call has mainly been studied in NCM (Vignal *et al.*, 2004a; Menardy *et al.*, 2012). In our study, CMM also appears to be a critical area for the neural representation of these calls: single units in CMM were both highly discriminant and highly selective for Distance calls and further studies are required to further explore the factors, such as experience, that might affect the representation of Distance calls in these two secondary areas.

Conclusion

We used a neuro-ethological approach to study the neural processes engaged in the perception of the meaningful information intrinsically coded in conspecific vocalizations. As underlined in recent reviews (Bennur *et al*., 2013; Theunissen & Elie, 2014), this approach might be the optimal way to understand the neural basis of discrimination, selectivity and invariance of neural responses in the auditory brain. With this study, we have also developed the birdsong model in a new direction: while the neuroethology of song production, learning and perception has been studied significantly, song birds had not yet been used to their full potential for studying the neural basis of vocal communication. As it has been stressed by Peter Marler (Marler, 2004b), the fact that birds have complex vocal repertoires and that their natural behavior is studied extensively make them perfect candidates for this endeavor. Our neurophysiological findings further justify this choice: we found that the majority of single neurons in the avian auditory cortex participate in the processing of semantic information, with a minority of neurons showing clear signs of semantic categorization. Although some of our results might be particular to the avian system, we also found general principles such as the correlation between discrimination, selectivity and invariance that will almost certainly be found in mammalian systems. Similarly, vertebrates might share some principles of neural computations for coding of semantic categories at the level of both the macro and micro circuitry. With future developments in both the avian and mammalian models, we believe that we will make great strides in understanding how vertebrates have resolved a common problem: understanding conspecific vocalizations (Kanwal & Rauschecker, 2007; Woolley & Portfors, 2013).

ACKNOWLEDGEMENT:

This work was supported by the NIH grant DC010132 to FET, a fellowship from the Fyssen Fondation to JEE and a CNRCNS NSF grant IIS1311446 to FET and JEE. The authors have no conflict of interest.

REFERENCES:

Amin, N., Doupe, A. & Theunissen, F.E. (2007) Development of selectivity for natural sounds in the songbird auditory forebrain. *J Neurophysiol*, **97**, 3517-3531.

Beckers, G.J. & Gahr, M. (2010) Neural processing of short-term recurrence in songbird vocal communication. *PLoS One*, **5**, e11129.

Beckers, G.J. & Gahr, M. (2012) Large-scale synchronized activity during vocal deviance detection in the zebra finch auditory forebrain. *Journal of Neuroscience*, **32**, 10594-10608.

Bennur, S., Tsunada, J., Cohen, Y.E. & Liu, R.C. (2013) Understanding the neurophysiological basis of auditory abilities for social communication: a perspective on the value of ethological paradigms. *Hear Res*, **305**, 3-9.

Bieszczad, K.M. & Weinberger, N.M. (2010) Remodeling the cortex in memory: Increased use of a learning strategy increases the representational area of relevant acoustic cues. *Neurobiology of learning and memory*, **94**, 127-144.

Bizley, J.K. & Cohen, Y.E. (2013) The what, where and how of auditory-object perception. *Nature reviews. Neuroscience*, **14**, 693-707.

Borst, A. & Theunissen, F.E. (1999) Information theory and neural coding. *Nat Neurosci*, **2**, 947-957.

Cohen, L. (1995) *Time-Frequency Analysis*. Prentice Hall, Englewood Cliffs, New Jersey.

Cohen, Y.E., Hauser, M.D. & Russ, B.E. (2006) Spontaneous processing of abstract categorical information in the ventrolateral prefrontal cortex. *Biology letters*, **2**, 261-265.

Cohen, Y.E., Russ, B.E., Davis, S.J., Baker, A.E., Ackelson, A.L. & Nitecki, R. (2009) A functional role for the ventrolateral prefrontal cortex in non-spatial auditory cognition. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 20045-20050.

David, S.V., Fritz, J.B. & Shamma, S.A. (2012) Task reward structure shapes rapid receptive field plasticity in auditory cortex. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, 2144-2149.

Elie, J.E., Mariette, M.M., Soula, H.A., Griffith, S.C., Mathevon, N. & Vignal, C. (2010) Vocal communication at the nest between mates in wild zebra finches: a private vocal duet? *Animal Behaviour*, **80**, 597-605.

Elie, J.E. & Theunissen, F. (in prep).

Gentner, T.Q. (2004) Neural systems for individual song recognition in adult birds. *Annals of the New York Academy of Sciences*, **1016**, 282-302.

Gentner, T.Q. & Margoliash, D. (2003) Neuronal populations and single cells representing learned auditory objects. *Nature*, **424**, 669-674.

Gifford, G.W., 3rd, Hauser, M.D. & Cohen, Y.E. (2003) Discrimination of functionally referential calls by laboratory-housed rhesus macaques: implications for neuroethological studies. *Brain, behavior and evolution*, **61**, 213-224.

Gifford, G.W., 3rd, MacLean, K.A., Hauser, M.D. & Cohen, Y.E. (2005a) The neurophysiology of functionally meaningful categories: macaque ventrolateral prefrontal cortex plays a critical role in spontaneous categorization of species-specific vocalizations. *J Cogn Neurosci.*, **17**, 1471-1482.

Gifford, G.W., 3rd, MacLean, K.A., Hauser, M.D. & Cohen, Y.E. (2005b) The neurophysiology of functionally meaningful categories: macaque ventrolateral prefrontal cortex plays a critical role in spontaneous categorization of species-specific vocalizations. *Journal of cognitive neuroscience*, **17**, 1471-1482.

Grace, J.A., Amin, N., Singh, N.C. & Theunissen, F.E. (2003) Selectivity for conspecific song in the zebra finch auditory forebrain. *J Neurophysiol*, **89**, 472-487.

Grimsley, J.M., Hazlett, E.G. & Wenstrup, J.J. (2013) Coding the meaning of sounds: contextual modulation of auditory responses in the basolateral amygdala. *Journal of Neuroscience*, **33**, 17538-17548.

Gunturkun, O. (2005) The avian 'prefrontal cortex' and cognition. *Curr Opin Neurobiol*, **15**, 686-693.

Gunturkun, O. (2012) The convergent evolution of neural substrates for cognition. *Psychological research*, **76**, 212-219.

Hauser, M. (1998) Functional referents and acoustic similarity: field playback experiments with rhesus monkeys. *Anim Behav*, **55**, 1647-1658.

Hsu, A., Woolley, S.M., Fremouw, T.E. & Theunissen, F.E. (2004a) Modulation power and phase spectrum of natural sounds enhance neural encoding performed by single auditory neurons. *Journal of Neuroscience*, **24**, 9201-9211.

Hsu, A., Woolley, S.M., Fremouw, T.E. & Theunissen, F.E. (2004b) Modulation power and phase spectrum of natural sounds enhance neural encoding performed by single auditory neurons. *Journal of Neuroscience*, **24**, 9201-9211.

Jeanne, J.M., Thompson, J.V., Sharpee, T.O. & Gentner, T.Q. (2011) Emergence of learned categorical representations within an auditory forebrain circuit. *Journal of Neuroscience*, **31**, 2595-2606.

Kanwal, J.S. & Rauschecker, J.P. (2007) Auditory cortex of bats and primates: managing species-specific calls for social communication. *Frontiers in bioscience : a journal and virtual library*, **12**, 4621-4640.

Machens, C.K., Schutze, H., Franz, A., Kolesnikova, O., Stemmler, M.B., Ronacher, B. & Herz, A.V. (2003) Single auditory neurons rapidly discriminate conspecific communication signals. *Nat Neurosci*, **6**, 341-342.

Manser, M.B. (2013) Semantic communication in vervet monkeys and other animals. *Animal Behaviour*, **86**, 491-496.

Marler, P. (2004a) Bird calls: their potential for behavioral neurobiology. *Ann N Y Acad Sci.*, **1016**, 31-44.

Marler, P. (2004b) Bird Calls: Their potential for neurobiology. In Zeigler, H.P., Marler, P. (eds) *Behavioral neurobiology of birdsong*. The New York Academy of Science, NY, pp. 31-44.

Meliza, C.D. & Margoliash, D. (2012) Emergence of selectivity and tolerance in the avian auditory cortex. *Journal of Neuroscience*, **32**, 15158-15168.

Menardy, F., Giret, N. & Del Negro, C. (2014) The presence of an audience modulates responses to familiar call stimuli in the male zebra finch forebrain. *The European journal of neuroscience*.

Menardy, F., Touiki, K., Dutrieux, G., Bozon, B., Vignal, C., Mathevon, N. & Del Negro, C. (2012) Social experience affects neuronal responses to male calls in adult female zebra finches. *The European journal of neuroscience*, **35**, 1322-1336.

Mouterde, S.C., Theunissen, F.E., Elie, J.E., Vignal, C. & Mathevon, N. (2014) Acoustic communication and sound degradation: how do the individual signatures of male and female zebra finch calls transmit over distance? *PLoS One*, **9**, e102842.

Nelken, I., Chechik, G., Mrsic-Flogel, T.D., King, A.J. & Schnupp, J.W. (2005) Encoding stimulus information by spike numbers and mean response time in primary auditory cortex. *J Comput Neurosci*, **19**, 199-221.

Nieder, A. (2012) Supramodal numerosity selectivity of neurons in primate prefrontal and posterior parietal cortices. *Proc Natl Acad Sci U S A*, **109**, 11860-11865.

Panzeri, S. & Treves, A. (1996) Analytical estimates of limited sampling biases in different information measures. *Network-Comp Neural*, **7**, 87-107.

Perez, E.C., Elie, J.E., Soulage, C.O., Soula, H.A., Mathevon, N. & Vignal, C. (2012) The acoustic expression of stress in a songbird: does corticosterone drive isolation-induced modifications of zebra finch calls? *Hormones and behavior*, **61**, 573-581.

Rauschecker, J.P. & Scott, S.K. (2009) Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nature Neuroscience*, **12**, 718-724.

Rauschecker, J.P. & Tian, B. (2000) Mechanisms and streams for processing of "what" and "where" in auditory cortex. *Proc Natl Acad Sci U S A*, **97**, 11800-11806.

Rieke, F., Bodnar, D.A. & Bialek, W. (1995) Naturalistic stimuli increase the rate and efficiency of information transmission by primary auditory afferents. *Proc R Soc Lond B Biol Sci*, **262**, 259-265.

Romanski, L.M. & Averbeck, B.B. (2009) The Primate Cortical Auditory System and Neural Representation of Conspecific Vocalizations. *Annual Review of Neuroscience*, **32**, 315-346.

Romanski, L.M., Averbeck, B.B. & Diltz, M. (2005) Neural representation of vocalizations in the primate ventrolateral prefrontal cortex. *J Neurophysiol*, **93**, 734-747.

Russ, B.E., Ackelson, A.L., Baker, A.E. & Cohen, Y.E. (2008) Coding of auditory-stimulus identity in the auditory non-spatial processing stream. *Journal of Neurophysiology*, **99**, 87-95.

Schnupp, J.W., Hall, T.M., Kokelaar, R.F. & Ahmed, B. (2006) Plasticity of temporal pattern codes for vocalization stimuli in primary auditory cortex. *Journal of Neuroscience*, **26**, 4785-4795.

Seyfarth, R.M. & Cheney, D.L. (2010) Production, usage, and comprehension in animal vocalizations. *Brain Lang*, **115**, 92-100.

Simpson, H.B. & Vicario, D.S. (1990) Brain pathways for learned and unlearned vocalizations differ in zebra finches. *Journal of Neuroscience*, **10**, 1541-1556.

Singh, N.C. & Theunissen, F.E. (2003) Modulation spectra of natural sounds and ethological theories of auditory processing. *J Acoust Soc Am*, **114**, 3394-3411.

Theunissen, F.E. & Doupe, A.J. (1998) Temporal and spectral sensitivity of complex auditory neurons in the nucleus HVc of male zebra finches. *Journal of Neuroscience*, **18**, 3786-3802.

Theunissen, F.E. & Elie, J.E. (2014) Neural processing of natural sounds. *Nature reviews. Neuroscience*, **15**, 355-366.

Theunissen, F.E., Sen, K. & Doupe, A.J. (2000) Spectral-temporal receptive fields of nonlinear auditory neurons obtained using natural sounds. *Journal of Neuroscience*, **20**, 2315-2331.

Tian, B., Reser, D., Durham, A., Kustov, A. & Rauschecker, J.P. (2001) Functional specialization in rhesus monkey auditory cortex. *Science*, **292**, 290-293.

Tsunada, J. & Cohen, Y.E. (2014) Neural mechanisms of auditory categorization: from across brain areas to within local microcircuits. *Front Neurosci*, **8**, 161.

Tsunada, J., Lee, J.H. & Cohen, Y.E. (2011) Representation of speech categories in the primate auditory cortex. *J Neurophysiol*, **105**, 2634-2646.

Tsunada, J., Lee, J.H. & Cohen, Y.E. (2012) Differential representation of auditory categories between cell classes in primate auditory cortex. *The Journal of physiology*, **590**, 3129-3139.

van Rossum, M.C.W. (2001) A novel spike distance. *Neural Computation*, **13**, 751-763.

Vicario, D.S., Naqvi, N.H. & Raksin, J.N. (2001) Behavioral discrimination of sexually dimorphic calls by male zebra finches requires an intact vocal motor pathway. *Journal of neurobiology*, **47**, 109-120.

Vignal, C., Attia, J., Mathevon, N. & Beauchaud, M. (2004a) Background noise does not modify song-induced genic activation in the bird brain. *Behav Brain Res*, **153**, 241-248.

Vignal, C., Mathevon, N. & Mottin, S. (2004b) Audience drives male songbird response to partner's voice. *Nature*, **430**, 448-451.

Wang, Y.A., Brzozowska-Prechtl & Karten, H.J. (2010) Laminar and columnar auditory cortex in avian brain. *Proc Natl Acad Sci U S A*, **107**, 12676-12681.

Woolley, S.M. & Portfors, C.V. (2013) Conserved mechanisms of vocalization coding in mammalian and songbird auditory midbrain. *Hear Res*, **305**, 45-56.

Zann, R. (1996) *The Zebra Finch: A Synthesis of Field and Laboratory Studies*. Oxford University Press, Oxford.

Figure 1. Four representative examples of neural responses to calls and song. In each panel (a-d) we show neural responses to 2 exemplars of 9 sound types (8 call types and song) for a representative neuron. The top row in each subpanel shows the spectrogram of the sound, the middle section the spike rasters obtained for 10 trials and the bottom row the PSTH. As explained in the methods, the sounds and responses are analysed in a 600 ms window but can start at any time point during that window. Note that here we show responses to 18 stimuli (9x2 examples) out of 130 that were played and analysed for these units. (A) Unit Selective for Wsst calls: this unit shows large responses to Wsst calls that peak towards the end of the call. (B) Unit Selective for Distance calls: this unit shows a strong patterned response to Distance calls. Weaker responses are also observed to Long Tonal calls, the precursor of the Distance call and to one song syllable. (C) Unit showing an intermediate level of selectivity: this unit responds the most to Wsst calls, Distance calls, and Distress calls with weaker responses to Thuk, Long Tonal calls and song syllables. Such unit might participate in a distributed code of semantic category found in an ensemble. (D) Non Selective Unit: this unit shows strong auditory responses to all sound types. Although this unit is non-selective, it responds to each stimulus with distinctive spike patterns and this information can also be used to categorize call types. Ne: Nest call; Te: Tet call; Th: Thuk call; Ws: Wsst call; DC: Distance call; Di: Distress call; Be: Begging call; LT: Long Tonal call; song: song Syllables.

Figure 2: Confusion Matrices and calculations of Inclusive and Exclusive Categorical Information

(A) Confusion matrix obtained at the end of the decoding procedure of the spike pattern responses of an example unit. The color in each bin of the matrix represents the joint probability that the neural responses to the stimulus i (rows: actual stimuli) were classified as belonging to the same stimulus (j predicted stimulus and here on the diagonal j=i) or the other stimuli (other columns in that row; j≠i). The Mutual Information (MI) measures the information content of the neural responses using these joint probabilities. MI_{max} is the theoretical upper bound of MI and depends on the stimulus set size.

(B) Inclusive-Categorical-uniform matrix obtained after a procedure that sets the joint probability outside the categories to equal and average values, effectively canceling out all information in the original confusion matrix (A) that did not pertain to the semantic classification and the classification of single calls within the categories. The mutual information of this modified matrix is called the Inclusive Categorical Information (ICI) of the unit (see methods). Note that the probabilities outside the block diagonal of this matrix are not zero but taken on very small values corresponding to very dark colors on our display.

(C) Exclusive-Categorical-uniform matrix obtained after a procedure that further removes the information in (B) about the classification of calls within categories by setting the joint probabilities within the block diagonal corresponding to each category to the same average value. The mutual information of that modified matrix is the Exclusive Categorical Information (ECI) of the unit. ECI_{max} is the theoretical upper bound of ECI and depends on the stimulus set size and composition.

Figure 3: Definition of Semantic Informative units in the avian auditory cortex: significance of the Exclusive Categorical Information.

(A) Histogram of the Exclusive Categorical Information (ECI) for confusion matrices obtained after random assignment of semantic labels to each predicted call stimulus of the confusion matrix obtained for a particular Semantic Informative unit. The actual value of ECI for the original confusion matrix of that unit is indicated by the purple line and is compared to the random distribution of ECI. Because the actual value of ECI is significantly higher (p<0.01) than what could be expected from the column-shuffled confusion matrices (see methods), the unit is labelled as "Semantic Informative", SI.

(B) Values of ECI for all single auditory units recorded (n=914). Units are classified into two populations: Semantic Informative units (SI, n=404) with significantly higher values of ECI (p<0.01) than expected by chance as defined in (A) and Semantic Uninformative units (SunI, n=510) with non-significant values of ECI. The number of units in each population is shown in the insert.

Figure 4: Proportion of information about semantic categories and individual vocalizations in the avian auditory cortex.

The three panels illustrate the values of total Mutual Information, MI (x-axes), Exclusive Categorical Information, ECI (y-axes in (A) and (B)), and Inclusive Categorical Information, ICI (y-axis in (C)), obtained for Semantic Informative (SI) single units and Semantic Uninformative (SunI) single units. The crosses in the (A) and (C) plots indicate the values of average ECI and ICI obtained by chance for the same units. To enable comparisons between values of information and units, values of MI, ECI and ICI of each unit were normalized by the theoretical upper bound of total MI (MI_{max}) of the stimulus dataset used for that unit in all plots. However, because the theoretical upper bound of ECI (ECI_{max}) is lower than MI_{max}, ECI values were also normalized by ECI_{max} in (B).

Figure 5: Semantic Informative units location in the avian auditory cortex. (A), (B) and (C) Left views of a 3D reconstruction of the positions of the single units recorded in the avian auditory cortex. (A) and (B) show 672 out of the 914 units that could be localized on anatomical slices while (C) only represents the 327 Semantic Informative units among the previous ones. The colour code indicates in (A) the auditory structures in which the position of the units could be identified; in (B) Semantic Informative (SI) versus Semantic Uninformative units as defined in Fig 3; in (C) the two populations of SI units that could be distinguished based on the dependence of their Exclusive Categorical Information to Acoustic and/or Semantic disruption as defined in Fig 6. The reference for the depth is the brain surface while it is the peak of the LPS (Lamina pallio-subpallialis) for the rostro-caudal axis. (D) Stacked bar plot of the proportion of the two populations of SI units (as defined in Fig 6) within each cortical auditory structure. Error bars are 95% confidence intervals on the proportion of SI units. Although SI units are found throughout the avian auditory cortex, their proportion is higher in Field L regions and in CMM than in CLM and NCM (X^2_5=17.5, p<0.01). The two populations of SI units are approximately evenly distributed between regions (X^2_{15}=10.1, p=0.07). L1, L2 and L3 are the three sub-regions of the avian primary auditory cortex Field L; NCM (Nidopallium caudomediale), CMM (Mesopallium caudomediale) and CLM (Mesopallium caudolaterale) are the sub-regions of the avian secondary auditory cortex; Unknown identify units that could be localized on slices but could not be unambiguously attributed to a particular zone.

Figure 6: Dependence of the Exclusive Categorical Information of Semantic Informative units on the acoustic correlation between stimuli and on the semantic labelling of stimuli.

(A) Effects of the disruption of the acoustic correlation within call categories (Acoustic disruption) and of the disruption of the semantic labelling within call categories (Semantic disruption) on the Exclusive Categorical Information (ECI) of the same example SI unit as Fig 3. Each point in the scatter plot indicates the value of ECI (y-axis left and middle plots or z-axis right plot), Semantic disruption (x-axis middle plot and y-axis right plot) and Acoustic disruption (x-axis left and right plots) obtained from shuffling different amounts (as a proportion) of calls across semantic categories (colour code) from the original confusion matrix. The value for the non-shuffled confusion matrix (the actual value of ECI) is shown with a red dot. Adjusted R-squares (R^2_a) quantify how well the acoustic correlation and the semantic labelling predict ECI. The ECI of this unit is very well predicted by semantic labelling and adding the acoustic correlation of calls as a predictor does not significantly increase the goodness-of-fit of the model (likelihood ratio test: F=0.28, p=0.76) while adding the semantic labelling of calls as a predictor to the model constructed only with the acoustic correlation of calls significantly increase the R^2_a by 0.03. This “A+S>A” unit is categorizing calls along semantic groups and this categorization cannot be fully explained by linear responses to spectro-temporal features in the calls.

(B) Histogram of the gain in adjusted R-squares of the model predicting the ECI of SI units when the Semantic disruption is added as a predictor in the model of ECI constructed with only the Acoustic correlation of calls. Most of SI units show significant increases of R^2_a (“A+S>A”): for these units, the semantic predictor provides additional information that is not explained by the acoustic predictor.

(C) Histogram of the gain in R^2_a of the model predicting the ECI of SI units when the Acoustic correlation of calls is added as a predictor in the model of ECI constructed with only Semantic disruption of the categories. For most of SI units the gain in R^2_a is null and non significant (“A+S=S”): for these units, the acoustic predictor does not provide any information that is not explained by the semantic predictor. The insert in (C) gives the number of SI units significant and non-significant for the two model comparisons.

Figure 7: Discrimination, selectivity and invariance for communication calls in the avian auditory cortex.

(A) Each scatterplot represents for one of the 9 call categories the values of discrimination (PCC, Percentage of Correct Classification), of selectivity (Sel, Selectivity index) and of invariance (Inv, Invariance Index) for the Semantic Informative (SI) units that are discriminating that category significantly better than chance. The insert in each scatterplot shows the number of these Discriminating SI units (purple bar) versus the number of non-discriminant SI units that are not shown on the scatterplot (pink bar). The inserts' y-axis maximum value is 300 units. The red horizontal lines are used to emphasize highly selective units for each call category in the sense that their discrimination for that category is greater than three times ('x3.36', Sel>21.75) their mean PCC for all other categories (see methods for the threshold choice). The green vertical bands show the average performance of the PCLDAS (Principal Component Discriminant Analysis on Spectrogram) for each semantic category (thickness: 2xSE, band centred on mean). Arrows in subplot of Wsst and Distance calls label the two units chosen as examples in Figure 8.

(B) Average percentage of correct classification (PCC) given by the PCLDAS (green bars) and by significantly discriminating SI units (purple bars) for each call category. Error bars are 2xSE. The PCC are significantly different between call categories (F_8=40.2, p<0.001) and between PCLDAS and units (F_1=1272.3, p<0.001). Significant effect of the interaction (F_8=28.8, p<0.001).

(C) Average Selectivity Index (Sel) for significantly discriminating SI units for each call category. Error bars are 2xSE. Sel is significantly different between categories (Kruskal-Wallis: X^2_8=532.6, p<0.001).

(D) Average Invariance Index (Inv) for Discriminating SI units for each call category. Error bars are 2xSE. Inv is significantly different between categories (Kruskal-Wallis: X^2_8=320.5, p<0.001).

Figure 8. Example of Responses of Selective Units with various level of invariance.

In each row, we show the confusion matrix (left) and a stimulus-response scatter plot (right) for a single unit that is classified as selective (Sel > 1.75) for Wsst calls (A) or Distance calls (B), and that has high level of invariance (A) or low level of invariance (B) for that category. The location, discrimination, selectivity and invariance properties of each unit for Wsst calls (A) or Distance calls (B) are indicated above each confusion matrix. In each row, the confusion matrix shows the conditional probability of decoding the stimulus identity using spike patterns. The stimulus-response curve shows the mean firing rate versus the projection of each stimulus spectrogram onto the acoustical feature dimension that would best discriminate a particular call type from all others. This acoustical feature is obtained in a logistic regression trained to distinguish a particular call type from all others based on spectrograms. Note that this regression is completely independent of the neural responses and was also performed using all sounds in our call database and not just the sounds used in each particular neural recording. The acoustical feature (i.e. also the coefficients of the logistic regression) is shown below the legend and to the right of the x-axis of the scatterplot, in the spectrogram space (x-axis limits 0-200 ms, y-axis limits 0-12 kHz). The acoustical features used for making these stimulus-response curves were obtained in the logistic regressions for Wsst calls (A) and Distance calls (B). The solid black line is the best fit in the minimum mean-square sense of a sigmoid function through the scatterplot. The sparse confusion matrices show that indeed these neurons are most selective for a single semantic category using a decoder that uses both spike counts and patterns. Note that the probability values in the block corresponding to Wsst calls are scattered for the more invariant unit (A) while probability values are concentrated on the diagonal in the block corresponding to Distance calls for the less invariant unit (B). The stimulus-response plots show that the selectivity can also clearly be seen in non-linear spike count responses to the presence of distinctive discriminative acoustical features. Ws: Wsst Call, Be: Begging Call, DC: Distance Call, Di:

Distress Call, LT: Long Tonal Call, Ne: Nest Call, So: Song Syllable, Te: Tet Call, Th: Thuk Call.

Figure 9: Discrimination, selectivity and invariance property differences between regions of the avian auditory cortex.

(A), (B) and (C) Each plot represents the average properties of Semantic Informative (SI) units of known location (n=327) in the 3 areas of the primary avian auditory cortex (Field L: L1, L2 and L3) and in the 3 areas of the secondary avian auditory cortex (CMM, CLM and NCM). Values are given as mean ± 2*SE. (A) Average over SI units of the mean percentage of correct classification (PCC) calculated over semantic categories. Field L regions have higher values of classification performances than secondary auditory areas (Effect of the 6 regions: Kruskal–Wallis or KW, X^2_5=26.0, p<0.001; comparison primary *vs* secondary regions: Mann–Whitney–Wilcoxon or MWW, U=12849, p<0.001). (B) Average values of Global Selectivity (GS) for SI units. GS is not different between regions (KW, X^2_5=3.5, p=0.63). (C) Average over SI units of the mean invariance (Inv), calculated over categories significantly discriminated. Invariance is mostly similar between regions (KW, X^2_5=10.4, p=0.07).

(D), (E) and (F) Each plot represents the average properties of SI units of known location for each semantic category and for each of the 6 auditory cortical regions. Semantic categories for which there is a significant region effect (KW test with p<0.01) are labelled in red. (D) Average percentage of correct classification (PCC) of all localized SI units (n=327). Wsst, Distress calls and Song syllables are better classified in Field L regions than secondary regions (Ws KW, X^2_5=32.8, p<0.001; Ws MWW, RS=12667, p<0.001; Di KW, X^2_5=18.2, p<0.01; Di MWW, RS=12672, p<0.001; So KW, X^2_5=30.5, p<0.001; So MWW, RS=13278, p<0.001), while Distance calls are better classified in L3 and CMM compare to other regions (KW, X^2_5=15.1, p<0.01). (E) Average Selectivity Index (Sel) for each semantic category over Discriminating SI units for that category. SI units significantly discriminating Distance calls are the most selective in CMM (KW, X^2_5=17.0, p<0.01). (F) Average Invariance value (Inv) for each semantic category over discriminating SI units. SI units significantly discriminating Wsst calls or Nest calls tend to be more invariant in L1 (Ws KW, X^2_5=11, p=0.05; Ne KW, X^2_5=13, p=0.02).

(G), (H) and (I) Left views of a 3D reconstruction of the positions of the 327 SI units correctly localized in the avian auditory cortex. The size of the balls represent in (G) the mean PCC of each unit, in (H) the Global Selectivity (GS) value for each unit and in (I) the mean Inv calculated over categories significantly discriminated for each unit. The colour represents in (G) and (I) the auditory structures in which the position of the units could be identified and in (H) the semantic category for which units are selective (Sel>1.75 or Sel>x3.36). Units that don't present a Sel value above 1.75 are considered as non-selective (see methods). The reference for the depth is the brain surface while it is the peak of the LPS (Lamina pallio-subpallialis) for the rostro-caudal axis. (G) Note that the size of the balls is bigger in the centre of the cloud of neurons compare to the periphery, indicating a core of units, mainly Field L units, with high discrimination performances for all semantic categories. (H) Note that units selective for Wsst calls, Distance calls and Song are everywhere in the auditory structures but that rostral units (mainly CMM) tend to be selective for Distance calls rather than other categories. (I) Invariance values are variable between units but evenly distributed between auditory structures.

Fig Supp1: Procedure used to generate sets of stimuli for each neurophysiological recording site.

The figure illustrates the three steps in the generation of one stimulus set for a given recording site: selection of the vocalizers among the ones available in the vocalization bank, selection of all the call types available for these vocalizers in the bank, random selection of 3 exemplars of calls for each call type of each vocalizer. Note that vocalizers were randomly selected at each site except for the 4 constant individuals that were always selected.

Fig Supp2: Correlation between the acoustic properties of the stimuli.

(A) Matrix of correlation values obtained between the spectrograms of the vocalizations used as stimuli in one neurophysiological recording site chosen as an example.

(B) Mean and standard deviation across stimulus sets (or equivalently across recording sites) of the average correlation values between the spectrograms of the vocalizations within the same category ("within category") and between the spectrograms of the vocalizations of each category and all the other spectrograms of the vocalizations in the stimulus set.

Fig Supp3: Selection of a threshold on Sel metric to consider a unit as selective.

Proportion of units that are selective (Sel>threshold) for at least one vocalization category (red dots) or for more than one vocalization category (blue dots) as a function of the Sel threshold tested. The threshold chosen on Sel to consider units as selective was 1.75.

Fig Supp4: Relationship between information measures and Discrimination, Selectivity and Invariance properties of Semantic Informative units.

(A), (B) and (C) represent with a colour code the Discrimination (A), Selectivity (B) and Invariance (C) of all the Semantic Informative units in the information space defined by the total Mutual Information of the confusion matrix (x-axes) and the Exclusive Categorical Information (ECI, y-axes).

(D), (E) and (F) represent with a colour code the Discrimination (D), Selectivity (E) and Invariance (F) of all the Semantic Informative units in the information space defined by the total Mutual Information of the confusion matrix (x-axes) and the Inclusive Categorical Information (ICI, y-axes).

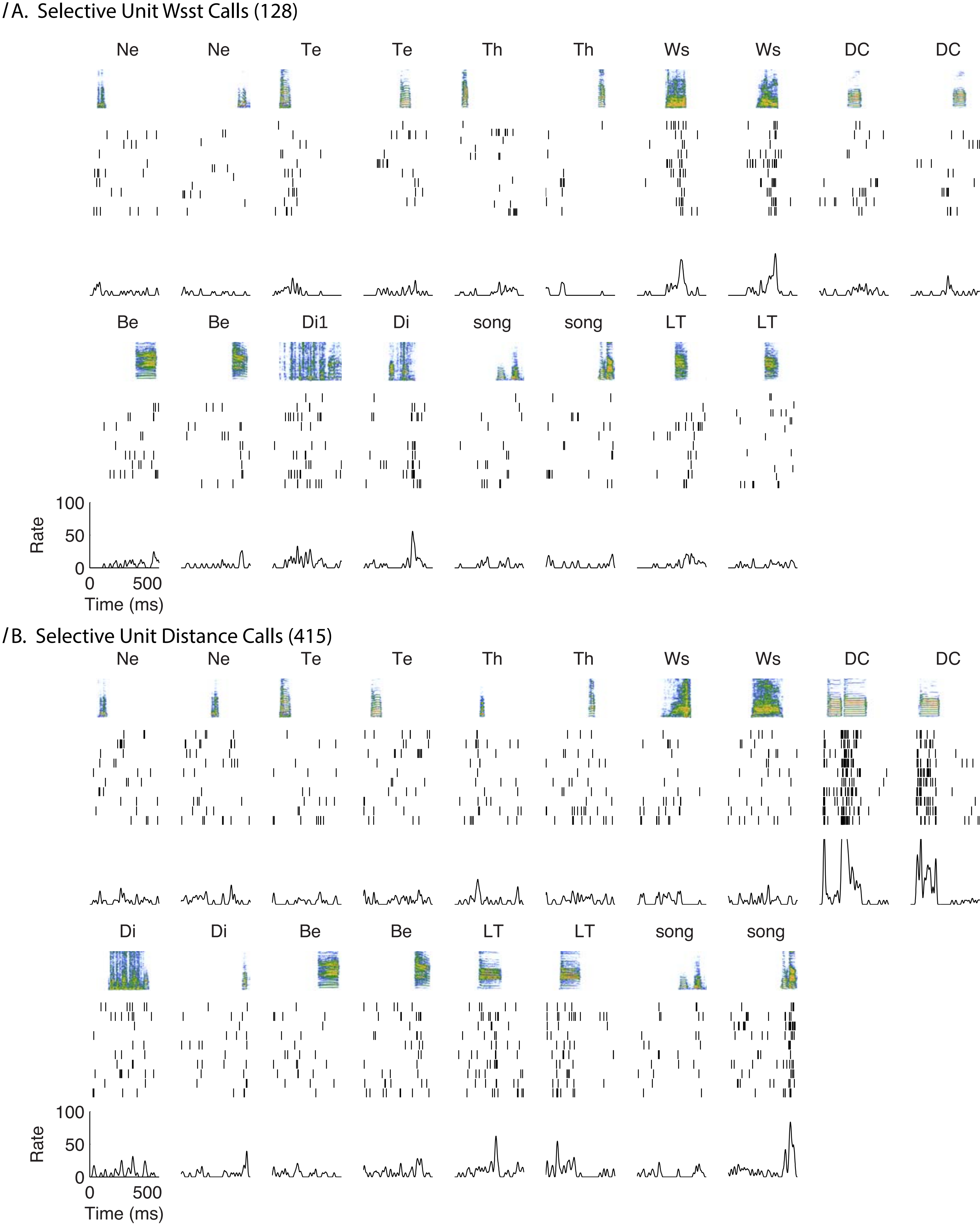
The average discrimination of units in (A) and (D) is measured by the average PCC over categories. Selectivity (B and E) is measured by the Global Selectivity index. Invariance (C and F) is measured for each unit by the average Inv value over the categories significantly discriminated.

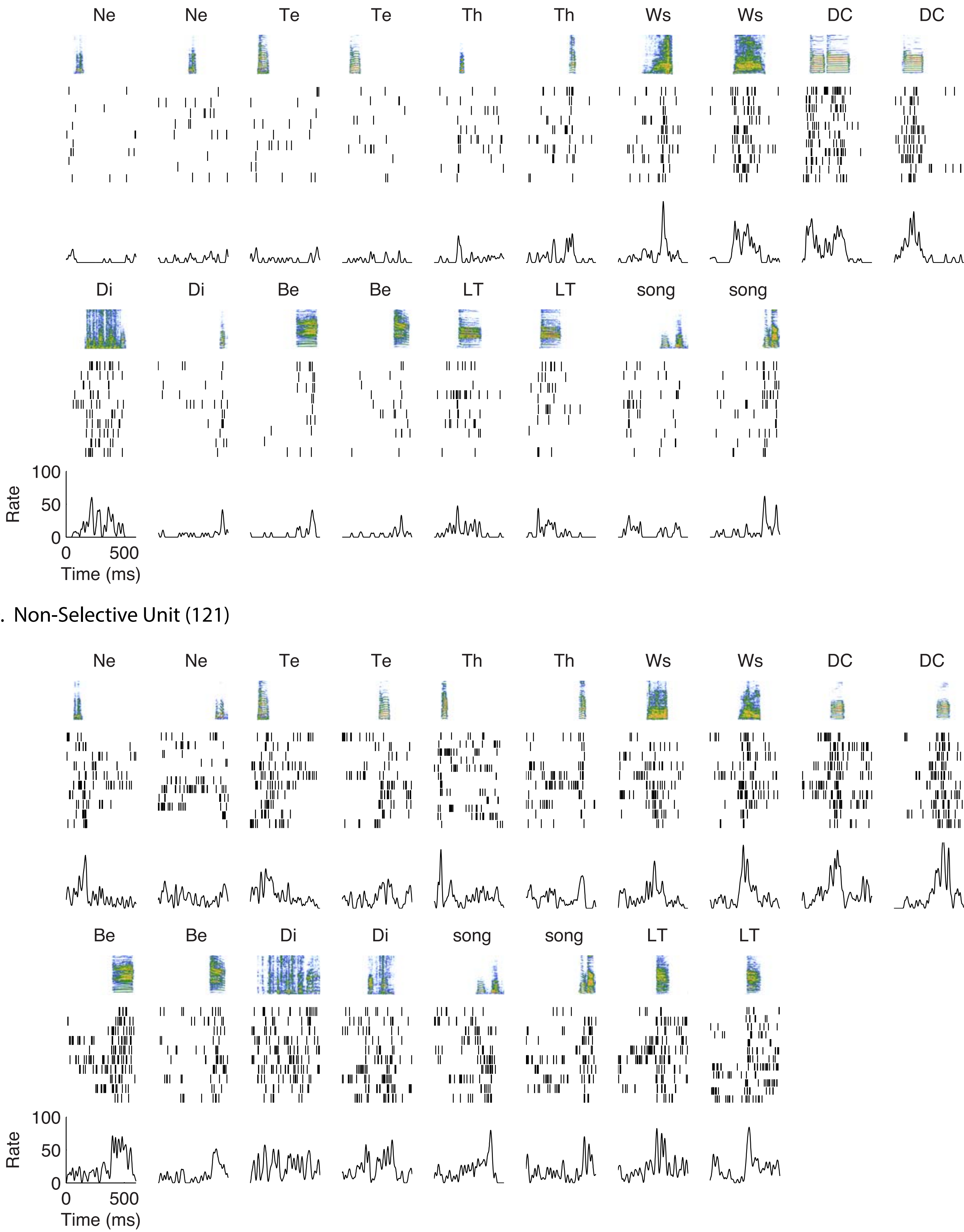
Category	All vocalizations from the bank					# vocalizations played back at each recording site (Mean±SD)
	Total number	# male vocalizations	# female vocalizations	# vocalizations per individual (Mean±SD)	# Individuals	
Wsst	30	18	12	2.5±0.9	12	9.4±3.6
Distance	114	60	54	8.1±1.6	14	13.6±4.7
Distress	16	7	9	1.8±1	9	4.8±2.6
Nest	96	45	51	8.7±0.9	11	12.1±3.9
Song	19	19		2.7±0.8	7	19.6±11.6
Tet	125	65	60	8.9±1	14	14.6±5.4
Thuk	37	14	23	6.2±1.9	6	6.5±2.3
Begging	24	12	12	3±0	8	9.1±3.6
Long tonal	25	15	10	6.25±3.8	4	6±2

Table 1 Constitution of the vocalization bank and of sets of vocalizations used during neurophysiological experiments

Semantic Category	PCC vs Sel		PCC vs Inv		Sel vs Inv	
	Rho	p	Rho	p	Rho	p
Wsst	0.72	<0.001	0.75	<0.001	0.64	<0.001
Begging	0.39	<0.001	0.37	<0.001	0.05	0.56
Distance	0.62	<0.001	0.63	<0.001	0.52	<0.001
Distress	0.78	<0.001	0.24	0.47	0.52	0.11
Long Tonal	0.66	<0.001	0.52	<0.001	0.33	<0.001
Nest	0.41	<0.001	0.43	<0.001	0.06	0.50
Tet	0.33	<0.001	0.42	<0.001	-0.06	0.49
Thuck	0.56	<0.001	0.52	<0.001	0.16	0.05
Song	0.38	<0.001	0.69	<0.001	0.31	<0.001

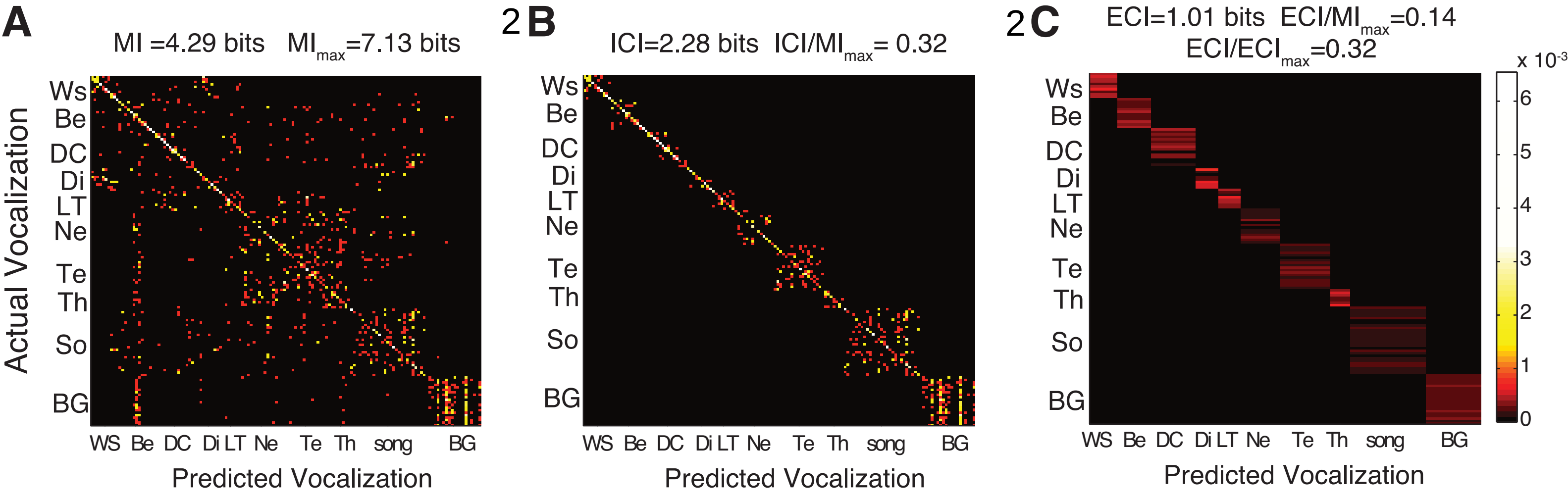
Table 2: Correlation between discrimination (PCC), selectivity (Sel) and invariance (Inv) properties of the units discriminating the semantic category above chance

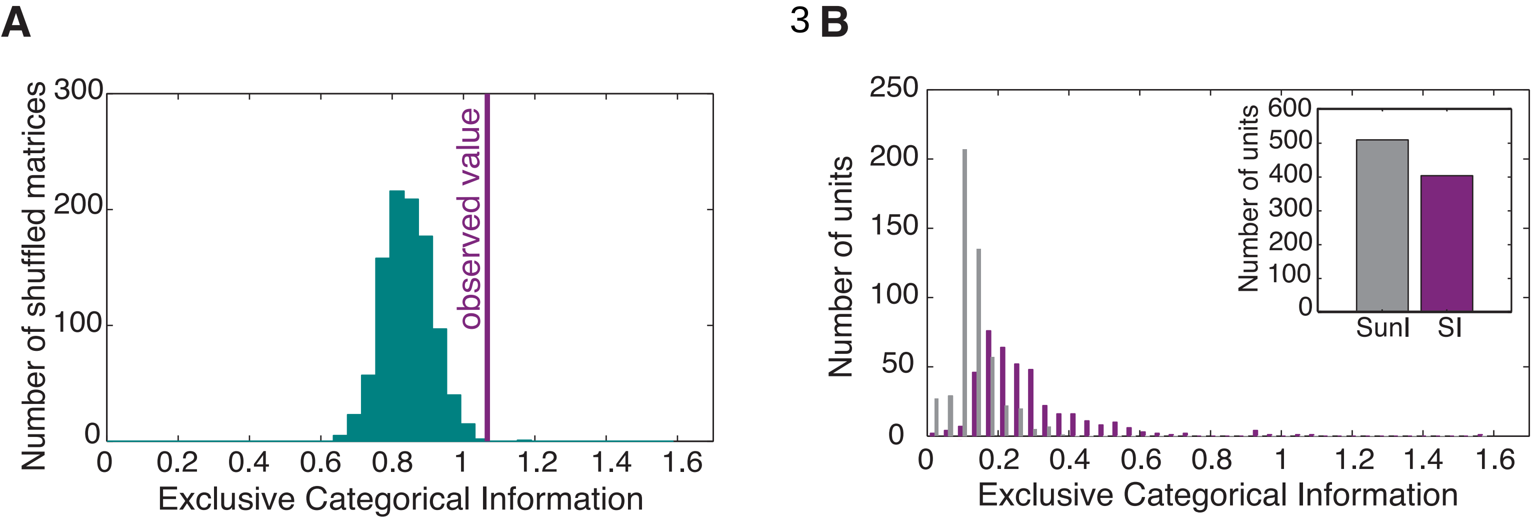

/ A. Selective Unit Wsst Calls (128)
Ne
Ne
Te
Te
Th
Th
Ws
Ws
DC
DC
Be
Be
Di1
Di
song
song
LT
LT
100
Rate
50
0
0
500
Time (ms)
/ B. Selective Unit Distance Calls (415)
Ne
Ne
Te
Te
Th
Th
Ws
Ws
DC
DC
Di
Di
Be
Be
LT
LT
song
song
100
Rate
50
0
0
500
Time (ms)

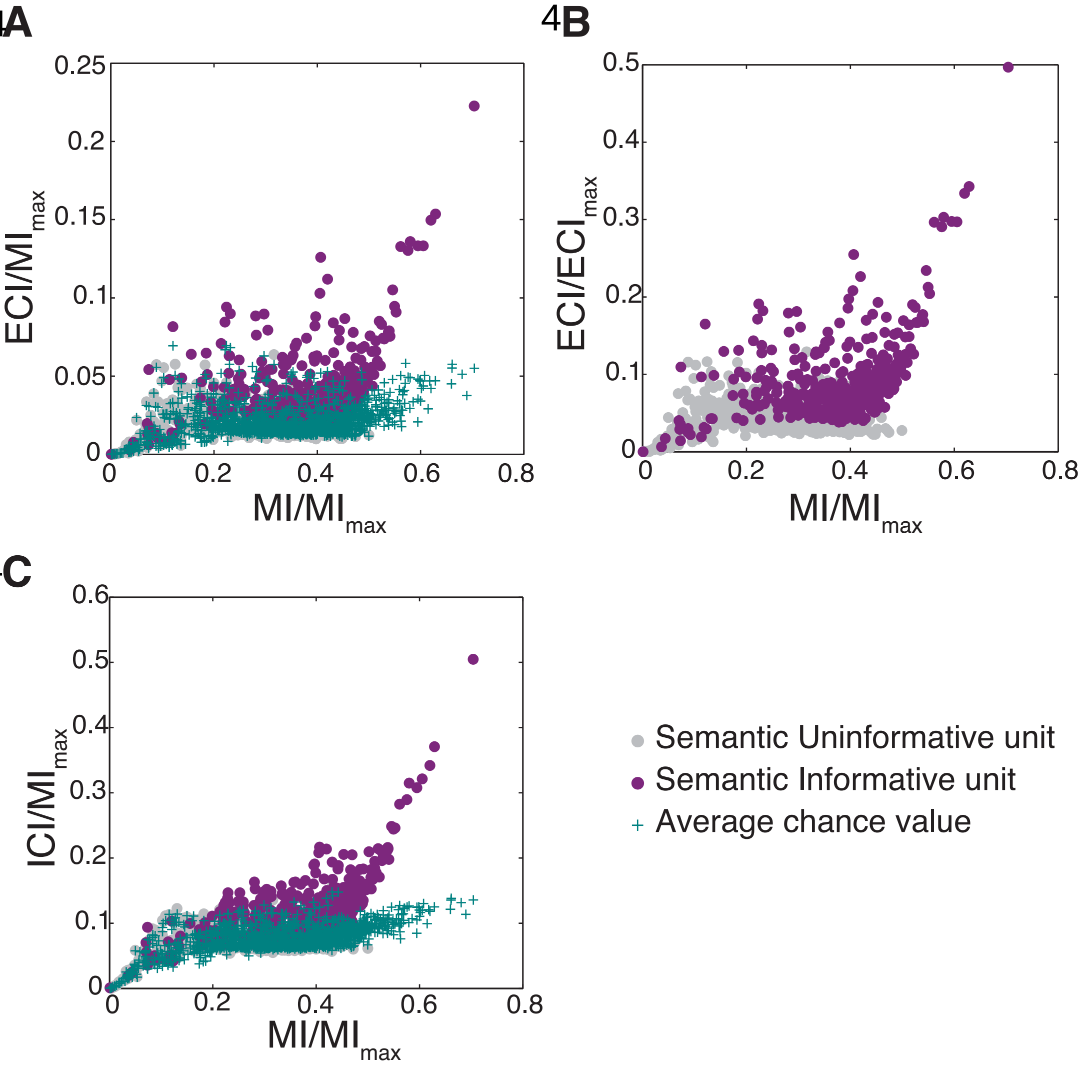


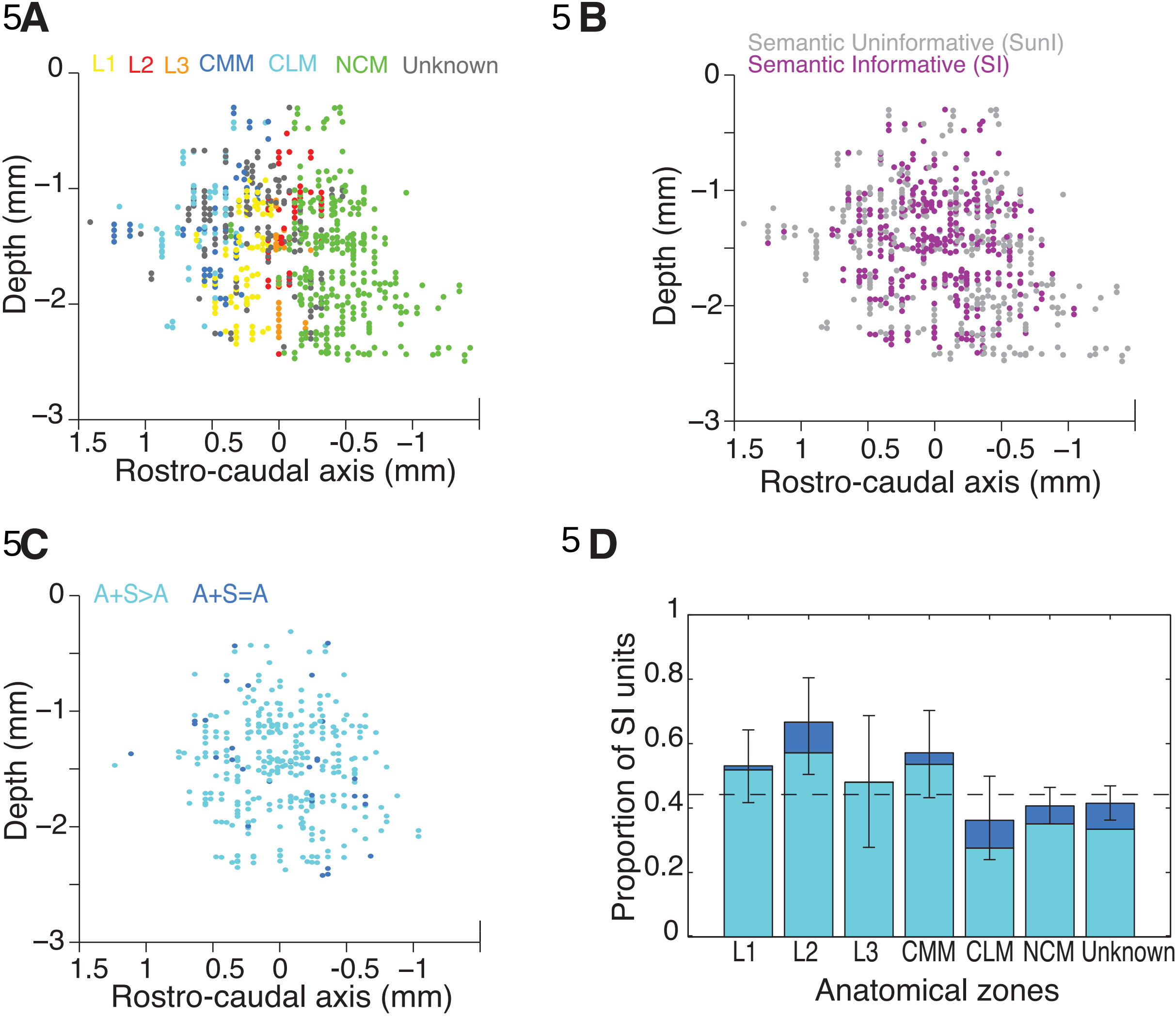
/C. Unit with Intermediate Selectivity (286)

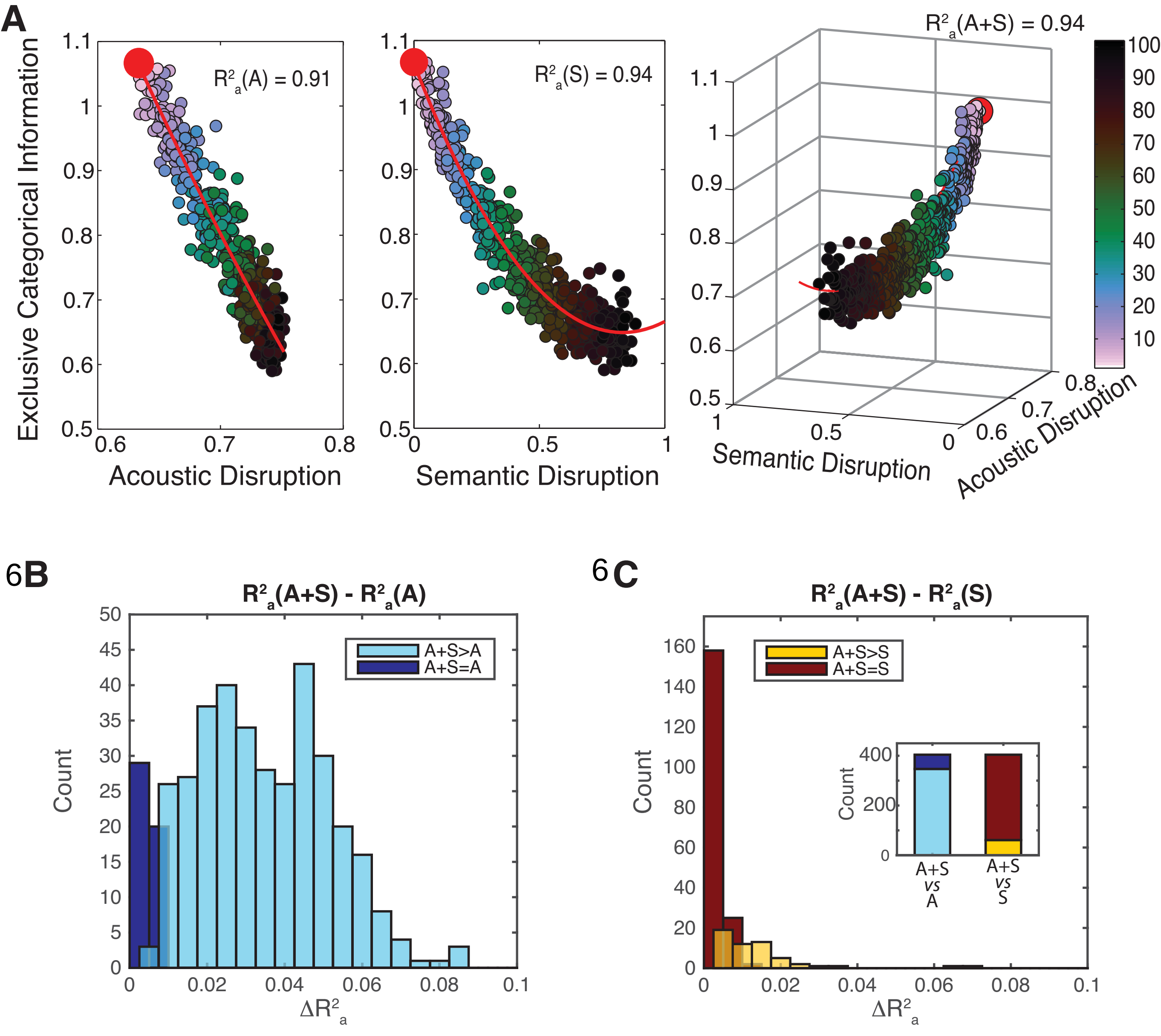
/D. Non-Selective Unit (121)

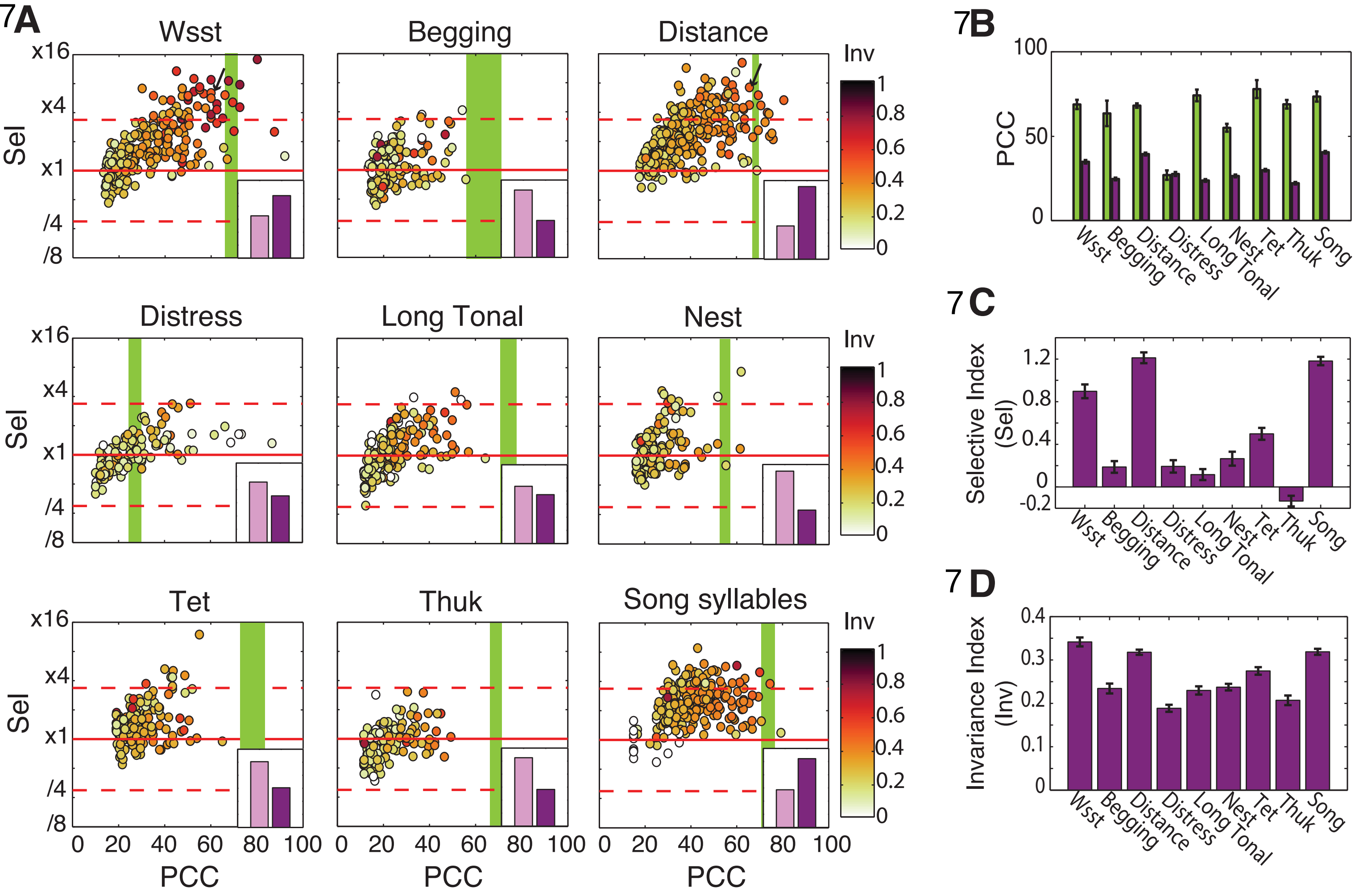

A
MI =4.29 bits MI$_{max}$=7.13 bits
Actual Vocalization
Ws
Be
DC
Di
LT
Ne
Te
Th
So
BG
WS Be DC Di LT Ne Te Th song BG
Predicted Vocalization
2 B
ICI=2.28 bits ICI/MI$_{max}$= 0.32
Ws
Be
DC
Di
LT
Ne
Te
Th
So
BG
WS Be DC Di LT Ne Te Th song BG
Predicted Vocalization
2 C
ECI=1.01 bits ECI/MI$_{max}$=0.14
ECI/ECI$_{max}$=0.32
Ws
Be
DC
Di
LT
Ne
Te
Th
So
BG
WS Be DC Di LT Ne Te Th song BG
Predicted Vocalization
x 10^{-3}
6
5
4
3
2
1
0

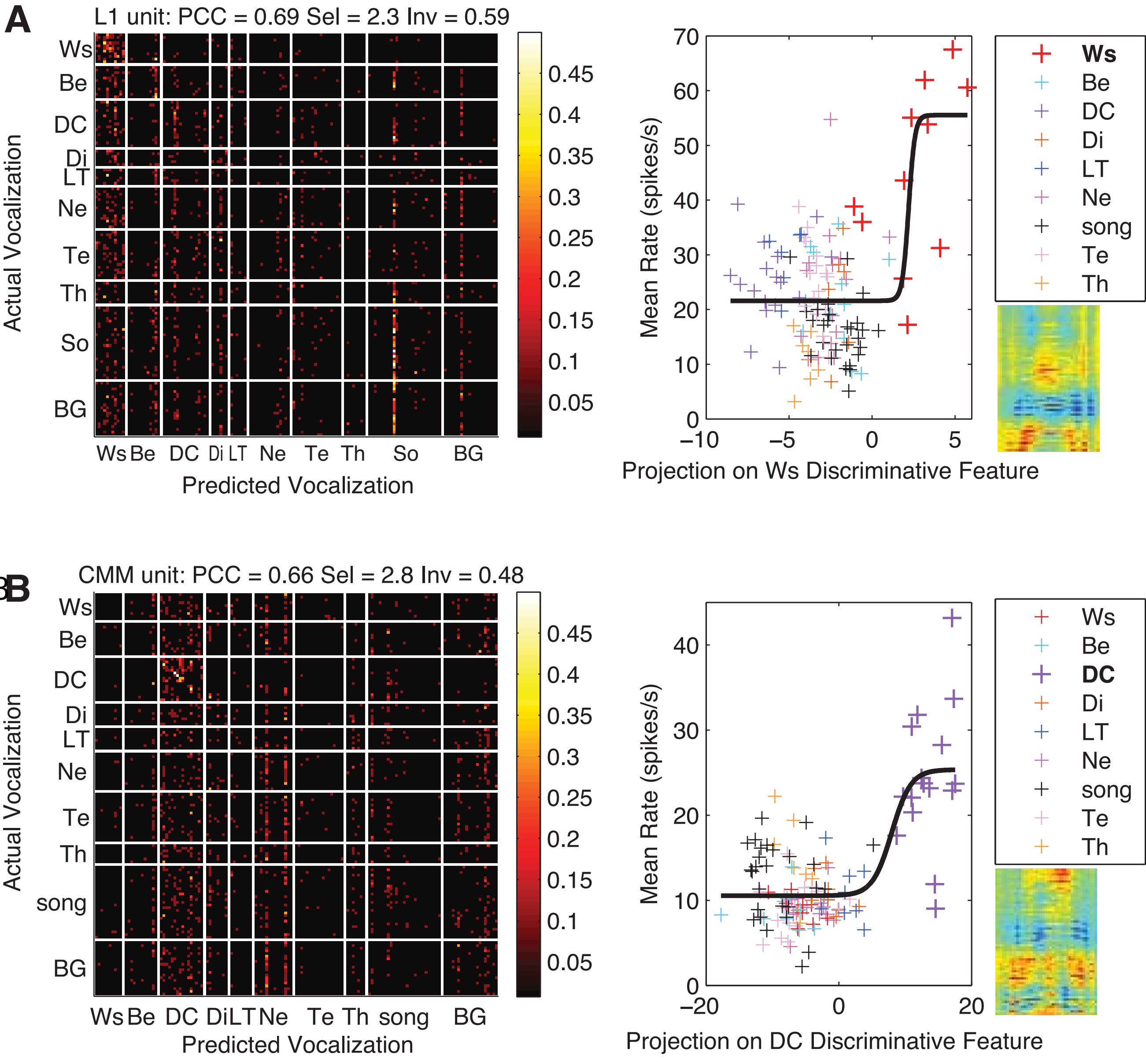

A
Number of shuffled matrices
300
200
100
0
observed value
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
Exclusive Categorical Information
3
B
Number of units
250
200
150
100
50
0
Number of units
600
500
400
300
200
100
0
SunI
SI
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
Exclusive Categorical Information

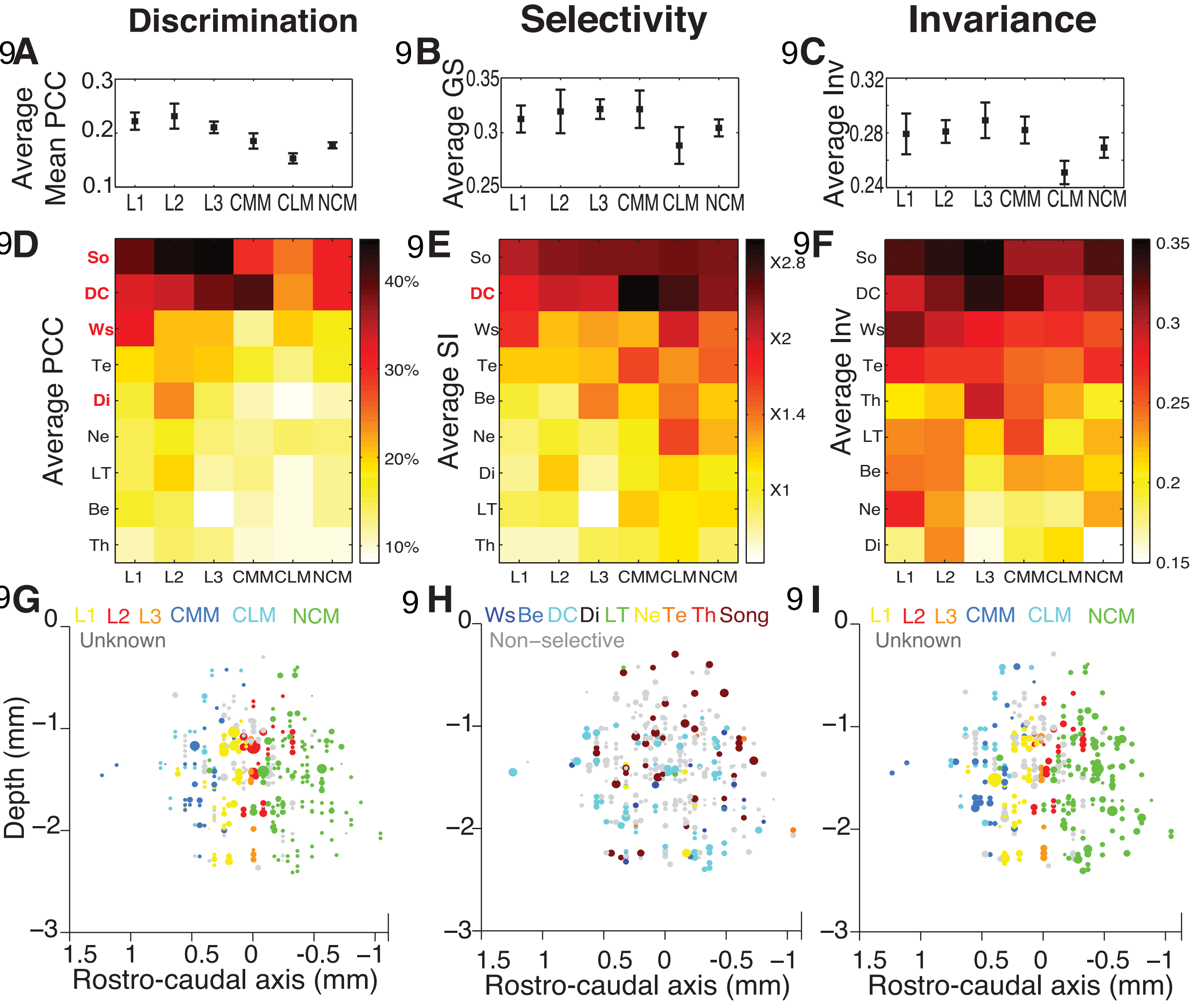

4A
0.25
0.2
0.15
0.1
0.05
0
ECI/MI$_{max}$
0
0.2
0.4
0.6
0.8
MI/MI$_{max}$
4B
0.5
0.4
0.3
0.2
0.1
0
ECI/ECI$_{max}$
0
0.2
0.4
0.6
0.8
MI/MI$_{max}$
4C
0.6
0.5
0.4
0.3
0.2
0.1
0
ICI/MI$_{max}$
0
0.2
0.4
0.6
0.8
MI/MI$_{max}$
● Semantic Uninformative unit
● Semantic Informative unit
+ Average chance value


5A
L1
L2
L3
CMM
CLM
NCM
Unknown
0
−1
−2
−3
Depth (mm)
1.5
1
0.5
0
-0.5
−1
Rostro-caudal axis (mm)
5 B
Semantic Uninformative (SunI)
Semantic Informative (SI)
0
−1
−2
−3
Depth (mm)
1.5
1
0.5
0
-0.5
−1
Rostro-caudal axis (mm)
5C
A+S>A
A+S=A
0
−1
−2
−3
Depth (mm)
1.5
1
0.5
0
-0.5
−1
Rostro-caudal axis (mm)
5 D
1
0.8
0.6
0.4
0.2
0
Proportion of SI units
L1
L2
L3
CMM
CLM
NCM
Unknown
Anatomical zones


A
Exclusive Categorical Information
$R^2_a(A) = 0.91$
Acoustic Disruption
0.6
0.7
0.8
0.5
0.6
0.7
0.8
0.9
1
1.1
$R^2_a(S) = 0.94$
Semantic Disruption
0
0.5
1
$R^2_a(A+S) = 0.94$
Semantic Disruption
Acoustic Disruption
0.6
0.7
0.8
100
90
80
70
60
50
40
30
20
10
6B
$R^2_a(A+S) - R^2_a(A)$
A+S>A
A+S=A
Count
ΔR^2_a
0
0.02
0.04
0.06
0.08
0.1
0
5
10
15
20
25
30
35
40
45
50
6C
$R^2_a(A+S) - R^2_a(S)$
A+S>S
A+S=S
Count
ΔR^2_a
0
20
40
60
80
100
120
140
160
Count
400
200
0
A+S vs A
A+S vs S


7A
Wsst
Begging
Distance
Distress
Long Tonal
Nest
Tet
Thuk
Song syllables
x16
x4
x1
/4
/8
Sel
Inv
1
0.8
0.6
0.4
0.2
0
0
20
40
60
80
100
PCC
7B
100
50
0
PCC
Wsst
Begging
Distance
Distress
Long Tonal
Nest
Tet
Thuk
Song
7C
Selective Index
(Sel)
1.2
0.8
0.4
0
-0.2
Wsst
Begging
Distance
Distress
Long Tonal
Nest
Tet
Thuk
Song
7D
Invariance Index
(Inv)
0.4
0.3
0.2
0.1
0
Wsst
Begging
Distance
Distress
Long Tonal
Nest
Tet
Thuk
Song


A
L1 unit: PCC = 0.69 Sel = 2.3 Inv = 0.59
Actual Vocalization
Ws
Be
DC
Di
LT
Ne
Te
Th
So
BG
Predicted Vocalization
Ws Be DC Di LT Ne Te Th So BG
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
Mean Rate (spikes/s)
0 10 20 30 40 50 60 70
−10 −5 0 5
Projection on Ws Discriminative Feature
Ws
Be
DC
Di
LT
Ne
song
Te
Th
B
CMM unit: PCC = 0.66 Sel = 2.8 Inv = 0.48
Actual Vocalization
Ws
Be
DC
Di
LT
Ne
Te
Th
song
BG
Predicted Vocalization
Ws Be DC Di LT Ne Te Th song BG
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
Mean Rate (spikes/s)
0 10 20 30 40
−20 0 20
Projection on DC Discriminative Feature
Ws
Be
DC
Di
LT
Ne
song
Te
Th


Discrimination
Selectivity
Invariance
9A
Average Mean PCC
0.3
0.2
0.1
L1
L2
L3
CMM
CLM
NCM
9B
Average GS
0.35
0.3
0.25
L1
L2
L3
CMM
CLM
NCM
9C
Average Inv
0.32
0.28
0.24
L1
L2
L3
CMM
CLM
NCM
9D
Average PCC
So
DC
Ws
Te
Di
Ne
LT
Be
Th
L1
L2
L3
CMM
CLM
NCM
40%
30%
20%
10%
9E
Average SI
So
DC
Ws
Te
Be
Ne
Di
LT
Th
L1
L2
L3
CMM
CLM
NCM
X2.8
X2
X1.4
X1
9F
Average Inv
So
DC
Ws
Te
Th
LT
Be
Ne
Di
L1
L2
L3
CMM
CLM
NCM
0.35
0.3
0.25
0.2
0.15
9G
L1 L2 L3 CMM CLM NCM
Unknown
Depth (mm)
0
−1
−2
−3
1.5
1
0.5
0
-0.5
−1
Rostro-caudal axis (mm)
9 H
Ws Be DC Di LT Ne Te Th Song
Non−selective
0
−1
−2
−3
1.5
1
0.5
0
-0.5
−1
Rostro-caudal axis (mm)
9 I
L1 L2 L3 CMM CLM NCM
Unknown
0
−1
−2
−3
1.5
1
0.5
0
-0.5
−1
Rostro-caudal axis (mm)

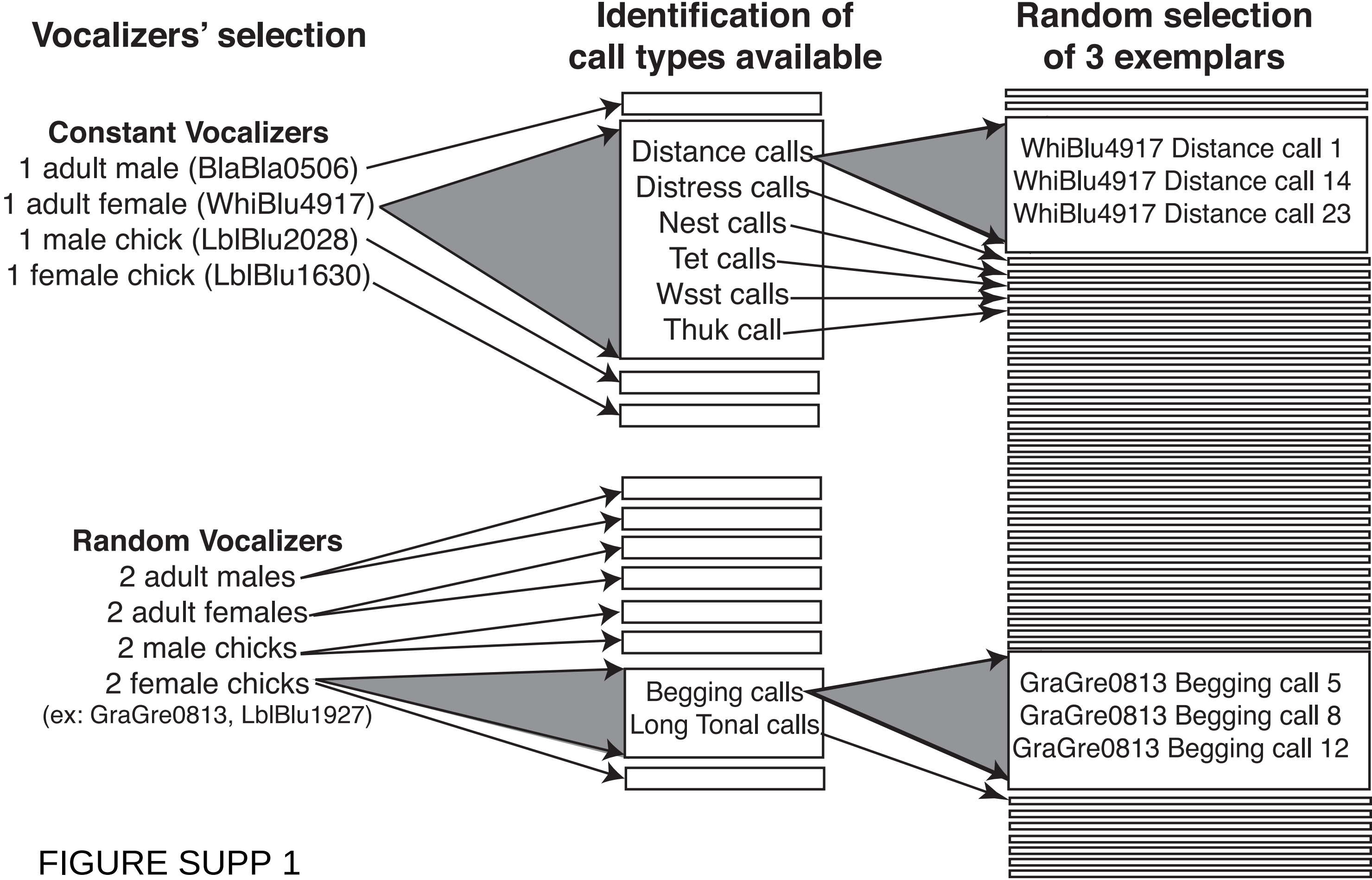
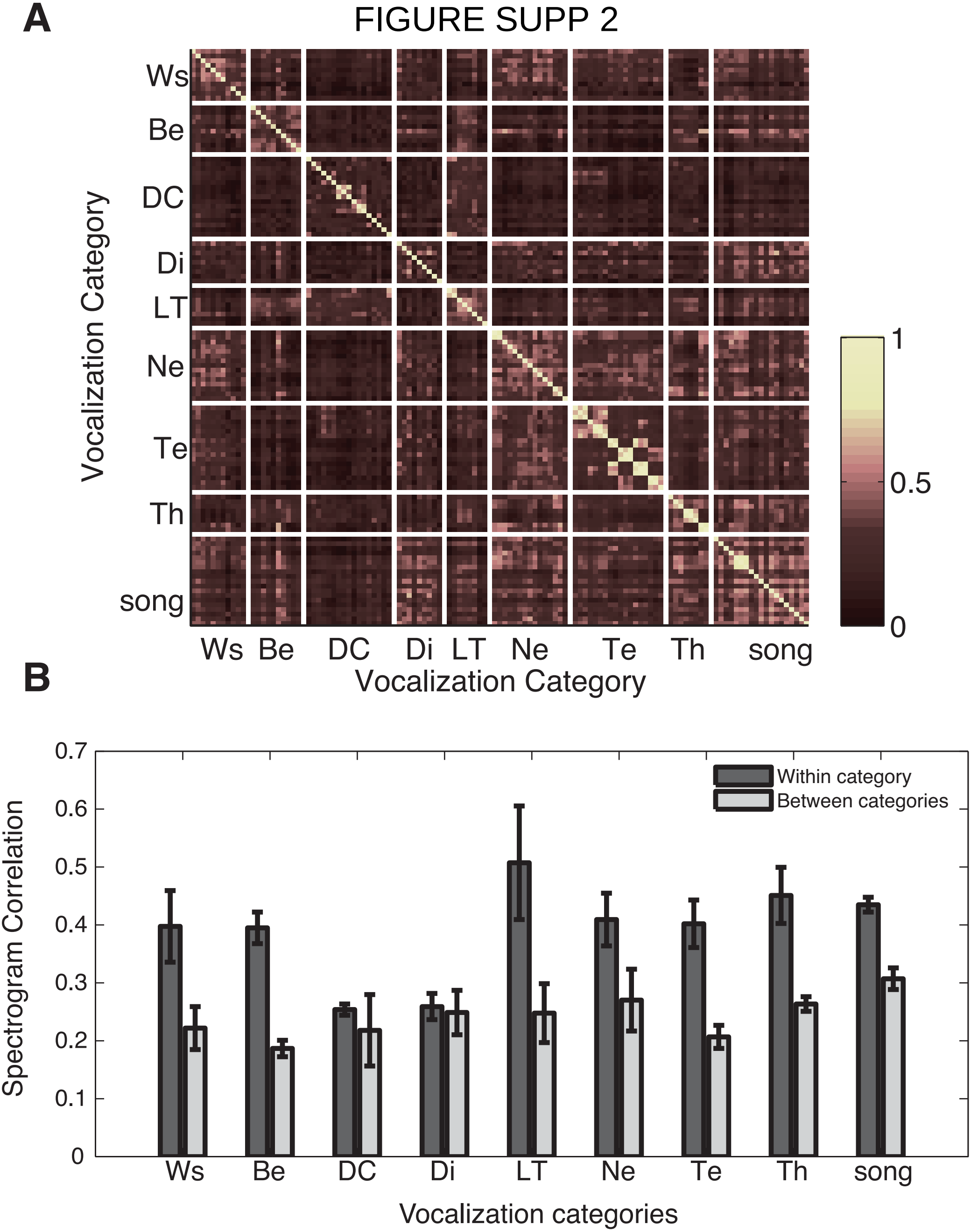

Vocalizers' selection
Identification of call types available
Random selection of 3 exemplars
Constant Vocalizers
1 adult male (BlaBla0506)
1 adult female (WhiBlu4917)
1 male chick (LblBlu2028)
1 female chick (LblBlu1630)
Distance calls
Distress calls
Nest calls
Tet calls
Wsst calls
Thuk call
WhiBlu4917 Distance call 1
WhiBlu4917 Distance call 14
WhiBlu4917 Distance call 23
Random Vocalizers
2 adult males
2 adult females
2 male chicks
2 female chicks
(ex: GraGre0813, LblBlu1927)
Begging calls
Long Tonal calls
GraGre0813 Begging call 5
GraGre0813 Begging call 8
GraGre0813 Begging call 12

FIGURE SUPP 1

FIGURE SUPP 2



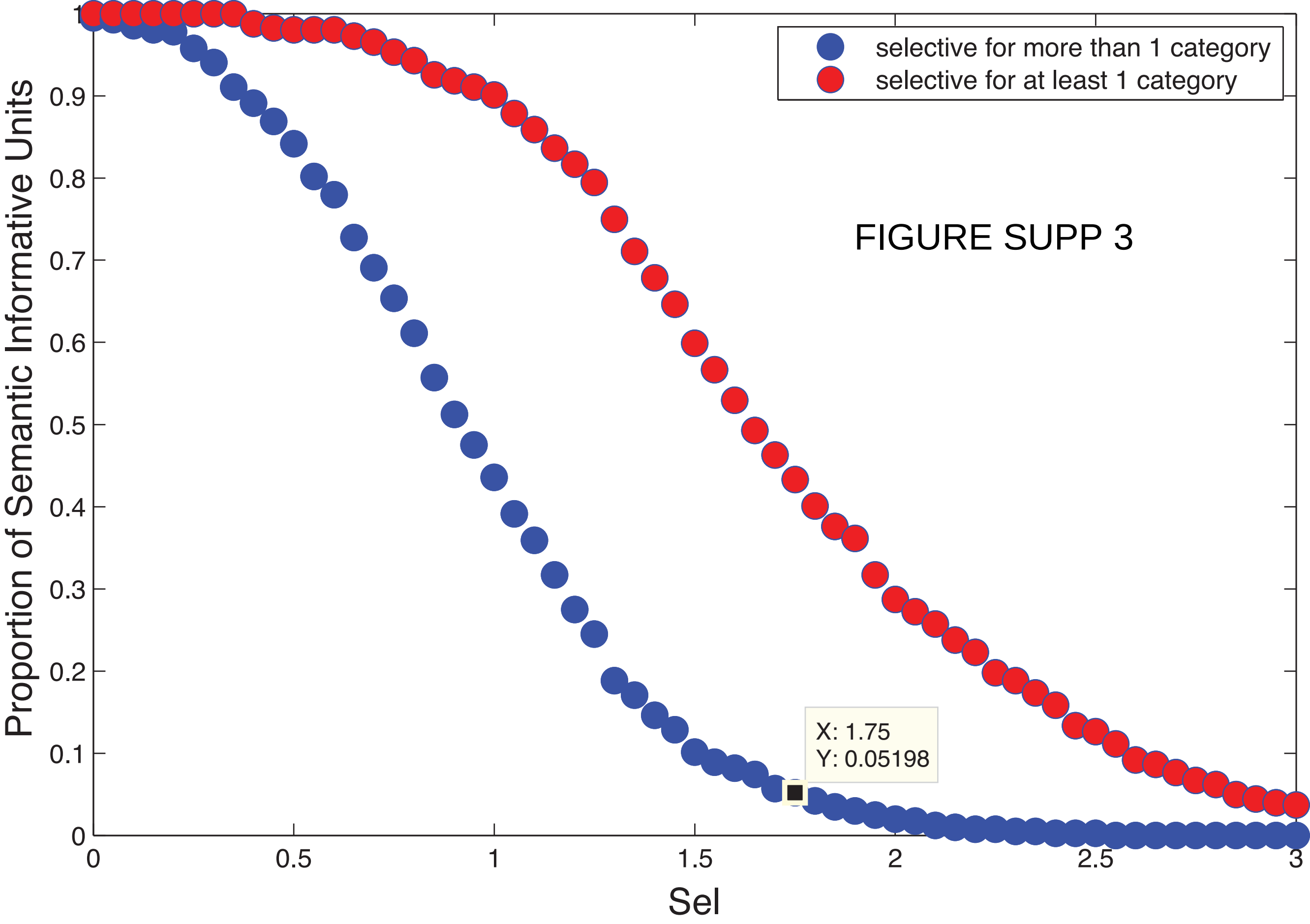
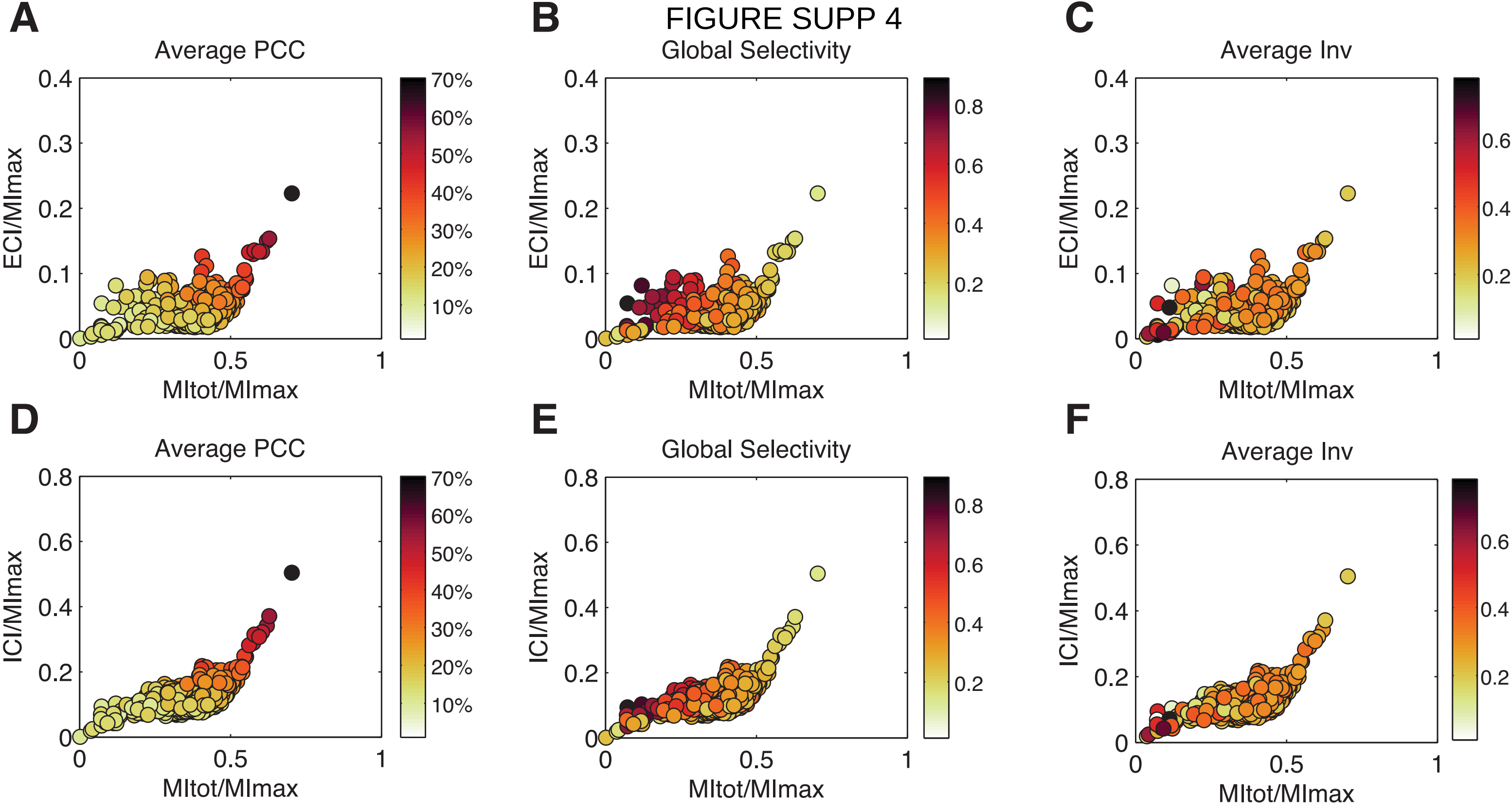

selective for more than 1 category
selective for at least 1 category
FIGURE SUPP 3
X: 1.75
Y: 0.05198
Proportion of Semantic Informative Units
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.5
1
1.5
2
2.5
3
Sel

FIGURE SUPP 4


A
Average PCC
ECI/MImax
MItot/MImax
B
Global Selectivity
ECI/MImax
MItot/MImax
C
Average Inv
ECI/MImax
MItot/MImax
D
Average PCC
ICI/MImax
MItot/MImax
E
Global Selectivity
ICI/MImax
MItot/MImax
F
Average Inv
ICI/MImax
MItot/MImax