

Research Article

Midbrain Synchrony to Envelope Structure Supports Behavioral Sensitivity to Single-Formant Vowel-Like Sounds in Noise

KENNETH S. HENRY,¹ KRISTINA S. ABRAMS,² JOHANNA FORST,¹ MATTHEW J. MENDER,¹
ERIKSON G. NEILANS,³ FABIO IDROBO,^{4,5} AND LAUREL H. CARNEY^{1,2}

¹*Department of Biomedical Engineering, University of Rochester, Rochester, NY 14642, USA*

²*Department of Neuroscience, University of Rochester, Rochester, NY 14642, USA*

³*Department of Psychology, University at Buffalo, Buffalo, NY 14260, USA*

⁴*Department of Psychological and Brain Sciences, Boston University, Boston, MA 02215, USA*

⁵*Universidad de Los Andes, Bogotá, Colombia*

Received: 1 August 2016; Accepted: 5 October 2016; Online publication: 20 October 2016

ABSTRACT

Vowels make a strong contribution to speech perception under natural conditions. Vowels are encoded in the auditory nerve primarily through neural synchrony to temporal fine structure and to envelope fluctuations rather than through average discharge rate. Neural synchrony is thought to contribute less to vowel coding in central auditory nuclei, consistent with more limited synchronization to fine structure and the emergence of average-rate coding of envelope fluctuations. However, this hypothesis is largely unexplored, especially in background noise. The present study examined coding mechanisms at the level of the midbrain that support behavioral sensitivity to simple vowel-like sounds using neurophysiological recordings and matched behavioral experiments in the budgerigar. Stimuli were harmonic tone complexes with energy concentrated at one spectral peak, or formant frequency, presented in quiet and in noise. Behavioral thresholds for formant-frequency discrimination decreased with increasing amplitude of stimulus envelope fluctuations, increased in noise, and were similar between budgerigars and humans. Multiunit recordings in awake birds showed that the midbrain encodes vowel-like sounds both through response synchrony to envelope structure and

through average rate. Whereas neural discrimination thresholds based on either coding scheme were sufficient to support behavioral thresholds in quiet, only synchrony-based neural thresholds could account for behavioral thresholds in background noise. These results reveal an incomplete transformation to average-rate coding of vowel-like sounds in the midbrain. Model simulations suggest that this transformation emerges due to modulation tuning, which is shared between birds and mammals. Furthermore, the results underscore the behavioral relevance of envelope synchrony in the midbrain for detection of small differences in vowel formant frequency under real-world listening conditions.

Keywords: average rate, budgerigar, formant frequency discrimination, fundamental frequency, inferior colliculus, synchronized rate

INTRODUCTION

Vowels are an important signal for the auditory system due to their strong contribution to speech intelligibility across languages (Ladefoged and Maddieson 1996; Kewley-Port et al. 2007). Whereas the fundamental frequency (F0) of vowels provides information about speaker gender and intonation, perception of contrasting vowels is based on differences in the shape of the frequency spectrum (Fig. 1a). Each vowel

Correspondence to: Kenneth S. Henry · Department of Biomedical Engineering · University of Rochester · Rochester, NY 14642, USA.
Telephone: (585) 276-3985; email: kenneth_henry@urmc.rochester.edu

contains energy concentrated at several spectral peaks, or formant frequencies, created by vocal tract filtering. The lowest two formant frequencies are typically sufficient for identification of vowels (Fant 1960; Hillenbrand et al. 1995). The neural mechanisms underlying auditory discrimination of vowels are poorly understood.

Behavioral experiments involving synthesized, vowel-like sounds show that humans can perceive small (0.5–2 %) changes in formant frequency (Kewley-Port and Watson 1994; Lyzenga and Horst 1995; Lyzenga and Horst 1997; Lyzenga and Horst 1998). Behavioral sensitivity is supported in the auditory nerve primarily by response synchrony to acoustic temporal structure rather than by average discharge rate. While auditory nerve fibers show differences in average rate across vowels at low sound levels that could potentially support behavioral discrimination (Sachs and Young 1980), this average-rate representation deteriorates at moderate-to-high sound levels and in noise due to rate saturation. In contrast, response synchrony to individual frequency components (Young and Sachs 1979; Sachs et al. 1983; Delgutte and Kiang 1984) and to F0-related envelope structure (Carney et al. 2015) provides a robust code for vowel discrimination across sound levels and in noise. Synchrony-based coding of vowels might be less effective in central nuclei, given lower-frequency synchronization limits compared to the auditory nerve (Joris et al. 2004). Central neurons may instead encode vowel formant structure through average discharge rate (Mesgarani et al. 2008; Perez et al. 2013; Carney et al. 2015; Honey and Schnupp 2015), but this hypothesis is untested in background noise. Here, we focus on coding of synthetic vowel-like sounds in the auditory midbrain.

The inferior colliculus (IC) is a large tonotopically organized nucleus in the midbrain and a nearly obligatory processing station in the ascending auditory pathway (Aitkin and Phillips 1984). Whereas the auditory nerve and nuclei of the brainstem encode amplitude-modulated (AM) sounds primarily through response synchrony to envelope structure (Joris and Yin 1992; Rhode and Greenberg 1994; Gleich and Klump 1995; Sayles et al. 2013), neurons at the level of the IC or higher encode AM through both envelope synchrony and substantial changes in average rate as a function of the modulation properties of the stimulus (IC: Langner and Schreiner, 1988; Woolley and Casseday, 2005; thalamus: Bartlett and Wang, 2007; cortex: Rosen et al. 2010; Yin et al. 2011; Johnson et al. 2012; reviewed in Joris et al. 2004). In commonly occurring IC neurons with band-enhanced modulation tuning, average rate increases with modulation depth for AM sounds presented within a limited band of modulation frequencies (Krishna and Semple 2000;

Nelson and Carney 2007). Given the prominent envelope structure of vowels, band-enhanced modulation tuning might also produce average-rate coding of vowels (Carney et al. 2015). This average-rate representation may be sensitive enough to support behavioral thresholds for formant-frequency discrimination. Alternatively, neural predictions of behavioral discrimination thresholds may require response synchrony to acoustic temporal structure in the IC.

The present study examined the neural mechanisms underlying formant-frequency discrimination of simple vowel-like sounds with energy concentrated at a single formant frequency using behavioral experiments and parallel neurophysiological recordings from the IC in the budgerigar. The budgerigar is a small parrot species and speech mimic with human-like behavioral sensitivity to complex sounds including AM (Dooling and Searcy 1981; Carney et al. 2013), consonants (Dooling et al. 1989), synthesized speech tokens along the /ra-la/ continuum (Dooling et al. 1995), and different vowels (Dooling and Brown 1990; reviewed in Dooling et al. 2000). Neurons in the budgerigar midbrain have band-enhanced modulation tuning, similar to that observed in mammals (Henry et al. 2016). Behavioral and neurophysiological experiments were conducted in quiet and in background noise. The results demonstrate an incomplete transformation to average-rate coding of simple vowel-like sounds in IC neurons with band-enhanced modulation tuning. Furthermore, the results underscore the significance of response synchrony to envelope structure in the IC for behavioral discrimination of vowel-like sounds under natural listening conditions.

MATERIALS AND METHODS

Behavioral and neurophysiological procedures in budgerigars (*Melopsittacus undulatus*) were approved by the University of Rochester Committee on Animal Resources. Experiments were conducted in English budgerigars, which are bred for larger size (40–65 g) and calmer deportment compared to other varieties of this species. Different birds were used for behavioral experiments (1 female, 3 males) and neurophysiological recordings (1 female, 2 males). Behavioral procedures in humans were approved by the Research Subjects Review Board of the University of Rochester. Human subjects (2 females, 1 male) ranged in age from 19 to 38 and had audiometric thresholds consistent with normal hearing (i.e., thresholds within 20 dB of 0-dB hearing level from 250 to 8000 Hz).

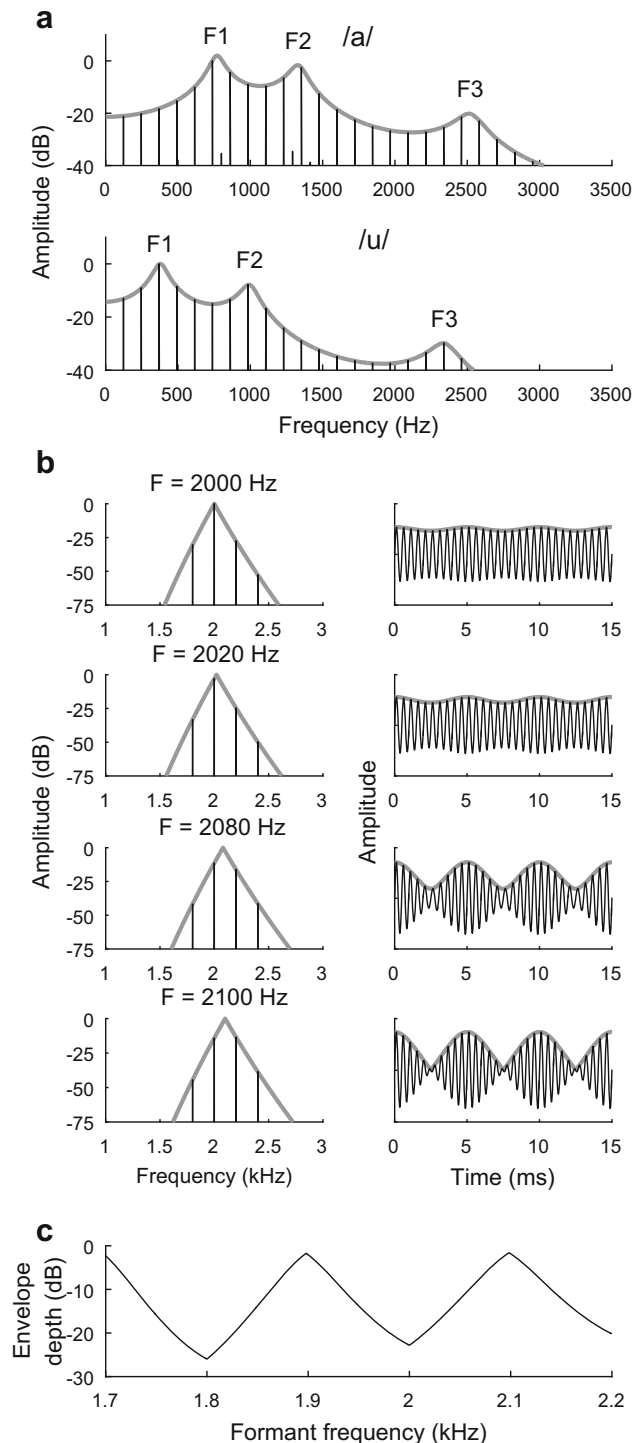


FIG. 1. Acoustic structure of vowels and simple vowel-like sounds with energy concentrated at a single formant frequency. **a** Amplitude spectra of synthetic vowels /a/ and /u/. The spectral envelope of each vowel (gray) exhibits formant peaks ($F1$ – $F3$; F = formant peak) due to resonant filtering of the laryngeal source signal. $F1$ and $F2$ are sufficient for vowel identification. Vowels were synthesized (Klatt and Klatt 1990) with fundamental frequency ($F0$) of 125 Hz and with four formant frequencies (Hillenbrand et al. 1995). **b** Single-formant stimuli used to study discrimination of formant frequency. Amplitude spectra (left) and waveforms (right) are plotted for stimuli with $F0$ of

200 Hz and formant frequencies ranging from 2000 Hz, which aligns with a harmonic (top), to 2100 Hz, which falls between two harmonics (bottom). Waveforms are plotted in arbitrary linear units with the temporal envelope (gray). Note the pronounced envelope fluctuations of the between-harmonic stimulus waveform. **c** Effective modulation depth of single-formant stimuli plotted as a function of formant frequency. Modulation depth was calculated as 20 times the base-10 logarithm of the quotient of the peak-to-peak amplitude of the temporal envelope over the mean of the temporal envelope

Behavioral Experiments

Single-formant stimuli were band-limited harmonic tone complexes with F0 of 200 Hz and a triangular spectral envelope (Fig. 1b), as used in previous human studies (Lyzenga and Horst 1995; Lyzenga and Horst 1997) and a modeling study of auditory nerve responses (Tan and Carney 2005). Frequency components were generated in sine phase and rolled off in amplitude by 200 dB per octave above and below the frequency of the formant. Only harmonics within 60 dB of the spectral peak were included in the stimulus. Stimuli were presented at 60 dB SPL with 25-ms \cos^2 onset and offset ramps and 250-ms duration.

Behavioral formant-frequency discrimination thresholds were estimated for two standard stimuli and for three background conditions using operant conditioning in trained budgerigars (six stimulus conditions total). The formant frequency of the standard stimulus was fixed at either 2000 Hz, which is aligned with a single harmonic, or 2100 Hz, which falls between two harmonics (Fig. 1b). Background conditions included quiet, steady-state noise, and fluctuating noise. Noise waveforms were generated independently for each trial, filtered to match the long-term average spectrum of speech (Byrne 1994), and gated on and off with the stimulus. The overall level of the steady-state noise was 73 dB SPL. The noise spectrum peaked at 400–500 Hz and rolled off by 13.4 dB at 2000 Hz. Fluctuating noise was square-wave modulated with 100 % depth and a modulation frequency of 16 Hz, as in a previous human study (George et al. 2006). Gating functions and square-wave modulation, when used, were imposed on the noise waveforms after scaling the level.

The test apparatus and procedures used to estimate behavioral auditory thresholds in trained budgerigars have been described previously (Carney et al. 2013; Henry et al. 2016). Briefly, behavioral testing was conducted in a sound isolation chamber (0.3 m³ inside volume) lined with 6.7-cm thick acoustic foam. Birds perched in a small wire-mesh cage positioned centrally on the floor of the chamber. Birds had access to three horizontally placed piezoelectric switches and the delivery tube of a seed dispenser. Acoustic stimuli were presented through a loudspeaker (Polk Audio MC60, Baltimore, MD USA) mounted 20 cm above the switches. Stimulus generation (50-kHz sampling frequency) and behavioral response acquisition were conducted with a data acquisition card (PCI-6151; National Instruments Corporation, Austin, TX USA) and controlled by custom software written in MATLAB (The MathWorks, Natick, MA USA). Stimuli were convolved with a digital pre-emphasis filter that compensated for the frequency

response of the system, followed by digital-to-analog conversion, power amplification (Crow D-75 A, Elkhart, IN USA), and presentation from the loudspeaker. The frequency response of the system was determined from the output of a calibrated microphone (Brüel and Kjær Type 4134; Marlborough, MA USA) placed inside the cage at the location of the animal's head. Tones were presented during calibration at 249 log-spaced frequencies from 0.050 to 15.1 kHz.

Behavioral testing was conducted using a single-interval, two-alternative, non-forced choice task and two-down, one-up, adaptive tracking procedures (Levitt 1970). Birds started each trial by pecking the center observing switch, which initiated presentation of either a standard stimulus, with formant frequency fixed at 2000 or 2100 Hz, or a target stimulus. The initial formant frequency of the target stimulus was 2100 Hz for the 2000-Hz standard and 2000 Hz for the 2100-Hz standard. Birds were trained to make a reporting response by pecking the left or right switch in response to each stimulus. For the 2000-Hz standard, the correct response to the standard was left and the correct response to the target was right. For the 2100-Hz standard, the correct response to the standard was right and the correct response to the target was left. This difference minimized retraining time between stimulus conditions by ensuring that for both standard stimuli, the correct response to the more modulated stimulus (i.e., with formant frequency closer to 2100 Hz; Fig. 1b) was on the same side. Reporting responses resulted in immediate termination of the stimulus. Correct responses were reinforced with delivery of 1–2 hulled millet seeds (dependent on bias, see below). Incorrect responses were reinforced with a 5-s timeout period during which the lights in the chamber were turned off. Responses made during the timeout reset the timeout timer, thus extending the total duration of the timeout; extended timeouts prevented birds from making multiple observing responses as the end of the timeout period approached. In rare instances in which the bird did not respond within 3 s of stimulus onset, a shorter 2-s timeout was imposed before the next trial could begin. Every block of ten trials was a random sequence of five standard and five target stimuli.

Birds were initially trained to discriminate the target stimulus from the standard stimulus, that is, a 100-Hz difference in formant frequency. Following mastery of this basic task, the formant-frequency discrimination threshold was estimated through systematic variation of the target frequency using two-down, one-up tracks. During tracking, each pair of consecutive hits at the same target frequency was followed by a step down in the absolute frequency

difference between target and standard, while each miss was followed by a step up (Fig. 2a). Step size was equal to the greater of two values, $25/n$ or 0.5 Hz, where n was the number of steps accumulated since the beginning of the test session (Robbins and Monro 1951; Levitt 1970). Tracks were allowed to continue for a minimum of 15 reversals in target frequency until two stability criteria were met: (1) the absolute difference in mean frequency between the final four reversals and the four preceding reversals was less than 5 Hz, and (2) the standard deviation of the

frequency of the final eight reversals was less than 5 Hz. The behavioral formant-frequency discrimination threshold of the track was calculated as the absolute difference between the mean of the final eight reversal points and the standard frequency. Response bias was calculated within tracks as -0.5 times the sum of the Z-score of the hit rate and the Z-score of the false-alarm rate (Macmillan and Creelman 2005). Bias was controlled by increasing the proportion of two-seed reinforcements for correct responses on the side that was biased against. Tracking sessions during which cumulative absolute bias exceeded 0.3 were excluded from further analysis.

Behavioral tracking sessions in budgerigars were repeated four–six times each day on the same stimulus condition. A minimum of 13 tracks were collected, until the formant-frequency discrimination thresholds estimated for the final six tracks met two stability criteria: (1) the absolute difference between the mean of the final three thresholds and the mean of the preceding three thresholds was less than 5 Hz, and (2) the standard deviation of the final six threshold estimates was less than 10 Hz. The mean of the final six threshold estimates that met these criteria was taken as the behavioral threshold for the stimulus condition. Stimulus conditions were tested in different orders in different birds.

As in the budgerigar behavioral experiments, formant-frequency discrimination thresholds in humans were assessed using a single-interval, two-alternative, non-forced choice task and two-down, one-up adaptive tracking. Tracking sessions were conducted in a walk-in sound isolation booth with a touchscreen computer running custom software written in MATLAB. The touchscreen interface consisted of a central pushbutton for initiation of each trial, two horizontally placed pushbuttons for acquisition of responses, and a feedback window for reporting correct and incorrect responses to the test subject. Stimuli were presented diotically using calibrated TDH-30 headphones. The initial formant frequency of the target stimulus was 50 Hz from that of the standard stimulus, except in cases of the on-harmonic standard stimulus presented in either noise or fluctuating noise. For these conditions, which preliminary testing showed were associated with higher thresholds, the initial target frequency was 100 Hz from the standard frequency. Tracks had 15 reversals, with a step size of $20/n$ Hz for the first 7 reversals, where n is the number of steps accumulated since the beginning of the track, and a step size of 1 Hz thereafter. The threshold estimate for each track was the absolute difference between the mean of the final eight reversals and the standard frequency. Thresholds were measured for a minimum of eight tracks until no further reduction in threshold was evident. The

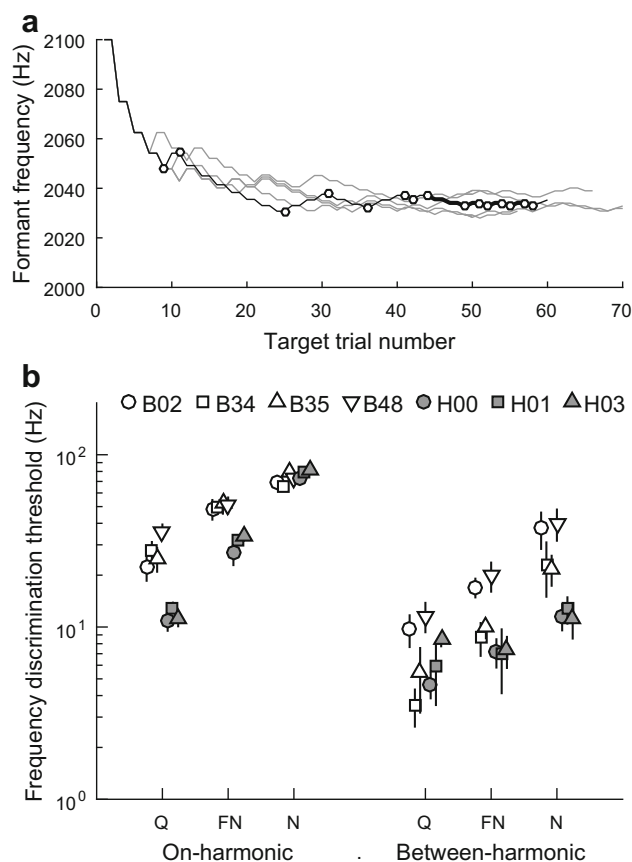


FIG. 2. Behavioral formant frequency discrimination in budgerigars and human subjects. **a** Representative data from behavioral tracking sessions in one budgerigar showing target formant frequency as a function of trial number. The frequency of the standard stimulus was 2000 Hz. The frequency of the target was adjusted from an initial value of 2100 Hz according to a two-down one-up tracking procedure (Levitt 1970) that converges on the behavioral discrimination threshold. Circles indicate reversal points in a representative track (B48–449; black trace) with threshold of 32.0 ± 2.7 Hz. **b** Mean formant frequency discrimination thresholds of four birds (open symbols) and three human subjects (gray symbols) plotted for the on-harmonic (2 kHz; left) and between-harmonic (2.1 kHz; right) standard stimulus in three background conditions: quiet (Q), fluctuating noise (FN), and unmodulated noise (N). Error bars indicate ± 1 SD from individual means. Discrimination thresholds in budgerigars and humans are lower (more sensitive) for between-harmonic than on-harmonic standards, and increase in continuous noise and, to a lesser extent, in fluctuating noise

mean of the final six thresholds with absolute cumulative bias < 0.3 was taken as the behavioral threshold for the stimulus condition. Stimulus conditions were tested in different orders in different subjects.

Neurophysiological Recording Procedures

Neurophysiological data were collected from the IC using chronically implanted microelectrodes in three awake, unrestrained birds (2 males; 1 female, implanted twice). The materials and methods used to implant electrodes and record from the budgerigar IC have been described previously (Henry et al. 2016). Briefly, tungsten microelectrodes (3–5 M Ω ; Microprobes, Gaithersburg, MD USA) were fixed to a miniature microdrive ('nDrive'; Neuronexus, Ann Arbor, MI USA) and lowered through a ~1-mm dorsal craniotomy site to the right IC under stereotaxic control in anesthetized birds. Anesthesia was induced with a subcutaneous injection of ketamine (3–5 mg/kg) and dexmedetomidine (0.08–0.1 mg/kg) and maintained throughout the 2–3-h implantation procedure by slow subcutaneous infusion with a syringe pump of ketamine (6–10 mg/kg/h), dexmedetomidine (0.16–0.27 mg/kg/h), and lactated Ringer's solution (30–50 ml/kg/h). Following the emergence of robust sound-evoked neural activity, the craniotomy was sealed with Kwik-Sil adhesive (World Precision Instruments, Sarasota, FL USA) and the base of the microdrive adhered to the skull with light-cured dental composite material (Kerr Vertise Flow; Orange, CA USA). Care was taken to position the electrode tip above the dorsal margin of the IC during the surgery, which allowed subsequent recordings from more ventral sites through adjustment of the microdrive control screw.

Daily neurophysiological recording sessions were conducted in awake, unrestrained birds beginning 1 week after the implantation procedure. The position of the recording site was held constant throughout the 2-h session, after which the electrode was advanced by 35 μ m. This distance reliably produced an increase in the best frequency (BF; the frequency of maximum sensitivity to tones) of the recording site. Recordings were conducted in a walk-in, double-walled sound isolation booth (8.27 m³ inside volume) lined with 6.7-mm thick acoustic foam. Birds perched in a small cage centered in the chamber and separated by 0.45 m in the horizontal plane from a loudspeaker (Dayton Audio PS180–8; Springboro, OH USA) facing the cage. Stimulus generation (50 kHz sampling frequency) and response acquisition (31.25 kHz sampling frequency) were conducted with a data acquisition card (National Instruments PCI-6251) and controlled by custom MATLAB software. Stimulus waveforms were calibrated with a digital preemphasis filter, as in the behavioral experiments,

before digital-to-analog conversion and power amplification (Tascam PA20 MK-II). Neural signals were buffered with a miniature custom headstage (operational amplifier-based voltage follower) located at the implant before passing through thin, flexible wires to a custom amplifier ($\times 1000$ – $\times 10,000$) that band-pass filtered the signal from 0.3 to 8 kHz.

Spikes were detected in multiunit neurophysiological recordings after high-pass filtering at 100 Hz (500-point FIR) to minimize the local field potential and transformation of the signal with the multiresolution Teager energy operator (Kim and Kim 2000; Choi et al. 2006). Representative transformed neural recordings are shown in a previous publication (see Fig. 2 of Henry et al. 2016). The response of the Teager energy operator is related to the instantaneous frequency and amplitude of the input signal and hence accentuates voltage spikes associated with neural action potentials. The amplitude threshold for action potential detection was set once per recording session at approximately half the peak amplitude of the largest peaks in the transformed neural signal (i.e., well above the level of the noise). Isolation of single-unit responses using a template-matching procedure was not possible because the action potential shape of neurons near the electrode tip were too similar for satisfactory discrimination. Nonetheless, multiunit recordings from the IC are expected to provide valuable insight into the response properties of this nucleus because previous studies have demonstrated robust anatomical gradients in both spectral and temporal response properties (Calford et al. 1985; Langner and Schreiner 1988; Langner et al. 2002; Baumann et al. 2011; but see Seshagiri and Delgutte 2007). Neighboring single neurons in cat IC, for example, have similar BF, frequency tuning bandwidth, and temporal integration statistics (Chen et al. 2012). Neighboring neurons in the budgerigar IC may also share similar response properties based on observations that all multiunit recording sites (1) have a single BF in response to tones and (2) show either band-enhanced modulation tuning with a single best modulation frequency (~90 % of sites) or high-pass modulation tuning (~10 %; see "Results"). More diverse tuning functions would be expected in the case of heterogeneous contributions from different individual neurons.

Formant Discrimination Thresholds of Individual Recording Sites

IC responses were recorded to single-formant stimuli presented at 60 dB SPL. F0 was set to 200 Hz, and formant frequencies ranged from 1900 to 2200 Hz (i.e., matched to the stimuli used in behavioral experiments). For recording sites tuned substantially

above or below 2000 Hz, additional stimuli were presented with formant frequencies spanning a 300-Hz range around the BF of the recording site. BF was estimated from responses to pure tones presented at 60 dB SPL. Formant frequencies were sampled with 20-Hz spacing for stimuli with the formant near a harmonic frequency and with a 3-Hz spacing for stimuli with the formant between two harmonics. This sampling resolution was adequate to capture fine variations in average discharge rate and synchronized rate as a function of formant frequency. Another subsample of recording sites was studied using both behavior- and BF-matched single formant stimuli presented in simultaneously gated, speech spectrum-shaped noise presented at 73 dB SPL. Noise waveforms were generated independently for each stimulus presentation. Stimuli were presented in random sequence for 20 repetitions with 25-ms \cos^2 onset and offset ramps, 250-ms duration, and 300-ms silent intervals between stimuli.

Neural formant-frequency discrimination thresholds were calculated based on both average discharge rate and synchronized rate to F0. Both response measures were calculated on a trial-by-trial basis over a 220-ms analysis window beginning 30 ms after stimulus onset. Synchronized discharge rate to F0, which reflects the magnitude of temporal oscillations in discharge rate synchronized to the envelope of the stimulus, was calculated as $\frac{1}{D} \sqrt{[\sum_{i=1}^n \cos(2\pi \cdot F_0 \cdot t_i)]^2 + [\sum_{i=1}^n \sin(2\pi \cdot F_0 \cdot t_i)]^2}$, where D is the duration of the analysis window in s, n is the total number of discharges, and t_i is the time of the i th discharge in s relative to stimulus onset. Neural discrimination thresholds were estimated using receiver operating characteristic (ROC) analysis (Egan 1975) of responses to formant frequencies spanning the 100-Hz range from 2000 to 2100 Hz and, for BF-matched stimuli, the 100-Hz range starting at the highest harmonic frequency below BF. ROC analysis computes the performance with which two stimuli can be classified based on a single observation of the response variable (average rate or synchronized rate), and hence provides a convenient framework for comparison to behavioral data. Formant-frequency discrimination thresholds were calculated as the minimum absolute change in formant frequency beyond which pairwise classification performance exceeded 70.7 % correct. This criterion corresponds to average behavioral performance at threshold for the two-down, one-up tracking procedure used in the behavioral experiments.

Formant-Frequency Discrimination Thresholds of the Pooled Neural Population

Formant-frequency discrimination thresholds of the pooled neural population were quantified based on average discharge rate and synchronized rate to F0

using a maximum likelihood-based pattern decoder analysis (Jazayeri and Movshon 2006; Day and Delgutte 2013). This decoder analysis calculates discrimination performance of the population based on single-trial, optimally weighted population responses and an assumption of independent, Poisson-distributed discharge counts [average rate analysis: total count; synchronized rate analysis: synchronized count; both response metrics showed a unity relationship between sample mean and bias-corrected variance (Gershon et al. 1998), consistent with a Poisson distribution]. Pairwise discrimination performance was calculated between each test frequency and the standard frequency (2000 or 2100 Hz) by first drawing 1000 population responses for each of the two stimulus alternatives at random. For each population draw, the logarithm of the likelihood of each alternative was calculated as in Jazayeri and Movshon (2006) as $\log L(\theta) = \sum_{i=1}^N n_i \log f_i(\theta) - \sum_{i=1}^N f_i(\theta) - \sum_{i=1}^N \log(n_i!)$, where n_i is the randomly drawn discharge count of the i th recording site, N is the total number of sites, and $f_i(\theta)$ is the mean discharge count of the i th site in response to stimulus condition θ . The first term is an optimally weighted sum of discharge counts across the population while the second term is the sum of average discharge counts. The last term can be ignored because it is independent of θ . For each random population draw, the selected discharge counts n_i were removed from the dataset prior to calculation of $f_i(\theta)$ to avoid overfitting the model. Discrimination performance was quantified as the proportion of population draws for which the log-likelihood of the correct stimulus condition was greater than that of the alternative. The population threshold was defined as the minimum change in frequency from that of the standard beyond which the classification performance of the decoder consistently exceeded 70.7 % correct. Population thresholds were estimated for behavior-matched stimuli presented in quiet and in simultaneously gated speech spectrum-shaped noise.

Models of the Auditory Nerve and Band-Enhanced IC Modulation Tuning

Computational models of the auditory nerve and band-enhanced modulating tuning in the IC were used to compare predicted responses to single-formant stimuli for these levels of the auditory pathway. The auditory nerve model (Zilany et al. 2014) has been rigorously tested against neurophysiological responses to a broad range of stimuli including tones, broadband noise, amplitude-modulated tones, and synthetic vowels. The model captures many of the non-linear response properties of auditory nerve fibers including compression, suppression, and increases in tuning bandwidth with

sound level. Furthermore, the model incorporates power law dynamics and long-term adaptation at the hair cell synapse that allow accurate prediction of responses to envelope periodicity and forward masking. The IC model (Mao and Carney 2015) generates band-enhanced average rate responses to AM stimuli through band-pass filtering of the time varying output of a single model auditory nerve fiber. Single-formant responses were predicted for the same band-limited harmonic tone complexes used in the behavioral experiments and neurophysiological recordings. Model auditory nerve responses were predicted for fibers with high spontaneous discharge rate, frequency tuning as in normal-hearing cat, and characteristic frequencies ranging from 1700 to 2400 Hz. Model IC responses were predicted with the band-pass filter centered on 200 Hz (i.e., best modulation frequency matched to F_0) and a Q value of 1.

RESULTS

Behavioral Thresholds for Formant Frequency Discrimination in the Budgerigar

Thresholds for behavioral frequency discrimination of single-formant stimuli were studied over approximately 400 behavioral test sessions in each bird using a single-interval, two-alternative, non-forced choice task and two-down, one-up, adaptive tracking (Fig. 2a). Formant discrimination thresholds in the budgerigar ($N = 4$) were lower (more sensitive) when the formant peak frequency of the standard stimulus was intermediate between two harmonics (2100 Hz) rather than aligned with a single harmonic (2000 Hz; Fig. 2b), coincident with stronger F_0 -related envelope fluctuations for between-harmonic stimuli (Tan and Carney 2005). Envelope fluctuations of between-harmonic stimuli are produced by beating between roughly equal-amplitude frequency components. The log amplitude of envelope fluctuations declines steadily as the formant frequency approaches on harmonic test frequencies (Fig. 1c). Formant-discrimination thresholds in simultaneously gated, speech-spectrum-shaped background noise were elevated compared to thresholds in quiet, with greater threshold elevation observed for steady noise than for fluctuating noise with periodic silent intervals.

Formant frequency discrimination thresholds in humans ($N = 3$) were assessed using matched stimuli and procedures (Fig. 2b, gray symbols). As in budgerigars, behavioral thresholds in humans showed an increase in sensitivity for the between-harmonic standard stimulus compared to the on-harmonic standard and release from masking in fluctuating noise compared to steady noise. Human thresholds overlapped with budgerigar thresholds for some stimulus conditions and were more sensitive for

others (most notably, for 2000 Hz standard stimulus in quiet and in fluctuating noise).

Frequency and Modulation Tuning Characteristics in the Budgerigar Midbrain

Neurophysiological recordings were obtained from the IC in three awake, unrestrained birds during daily recording sessions beginning 1 week after electrode implantation. The frequency and modulation tuning properties of recording sites in the budgerigar IC have been described previously (Henry et al. 2016). Briefly, individual recording sites ($N = 64$) showed excitatory rate responses to tone stimuli with band-limited frequency tuning (Fig. 3a, left). BF increased from approximately 400 Hz at dorsal recording sites to 5 kHz at the most ventral sites. The 10-dB bandwidth of the excitatory rate response to tones increased with increasing BF, while mean Q_{10} (BF divided by 10-dB bandwidth; Fig. 3a, right) increased from 2.0 at BF of 1 kHz to 3.25 at BF of 4 kHz (similar to cat IC over the same BF range; Ramachandran et al. 1999).

Neural recording sites in the IC usually showed band-enhanced modulation tuning in response to AM tone stimuli with carrier frequency equal to BF (Fig. 3b, top left; 60/64 sites); that is, AM tones evoked greater average discharge rate than did unmodulated tones over a limited band of modulation frequencies. The best modulation frequency (BMF) of each recording site was quantified as the modulation frequency that evoked the largest increase in discharge rate compared to the response to the unmodulated stimulus. BMFs ranged from 64 to 512 Hz across recording sites (median 161 Hz; Fig. 3b, top right). Rate enhancement, calculated as the percent difference in response rate between stimuli presented without modulation and stimuli presented at BMF with 100 % modulation depth, ranged from 28 to 185 % across recording sites (median 104 %; Fig. 3b, bottom). Maximal BMFs and rate enhancement were observed at recording sites with intermediate BFs (1.5–3.5 kHz). The recording sites without band-enhanced modulation tuning had rate-based modulation tuning that was high-pass in shape (4/64 sites).

Rate and Synchrony-Based Representations of Formant Frequency Can Account for Behavioral Discrimination Thresholds in Quiet

Neural recording sites in the IC typically showed robust excitatory rate responses to single-formant stimuli presented in quiet. The average discharge rate of individual recording sites varied non-monotonically with formant frequency in quiet (Fig. 4b, top). Local peaks in average rate were not strongly related to the proximity of the formant peak

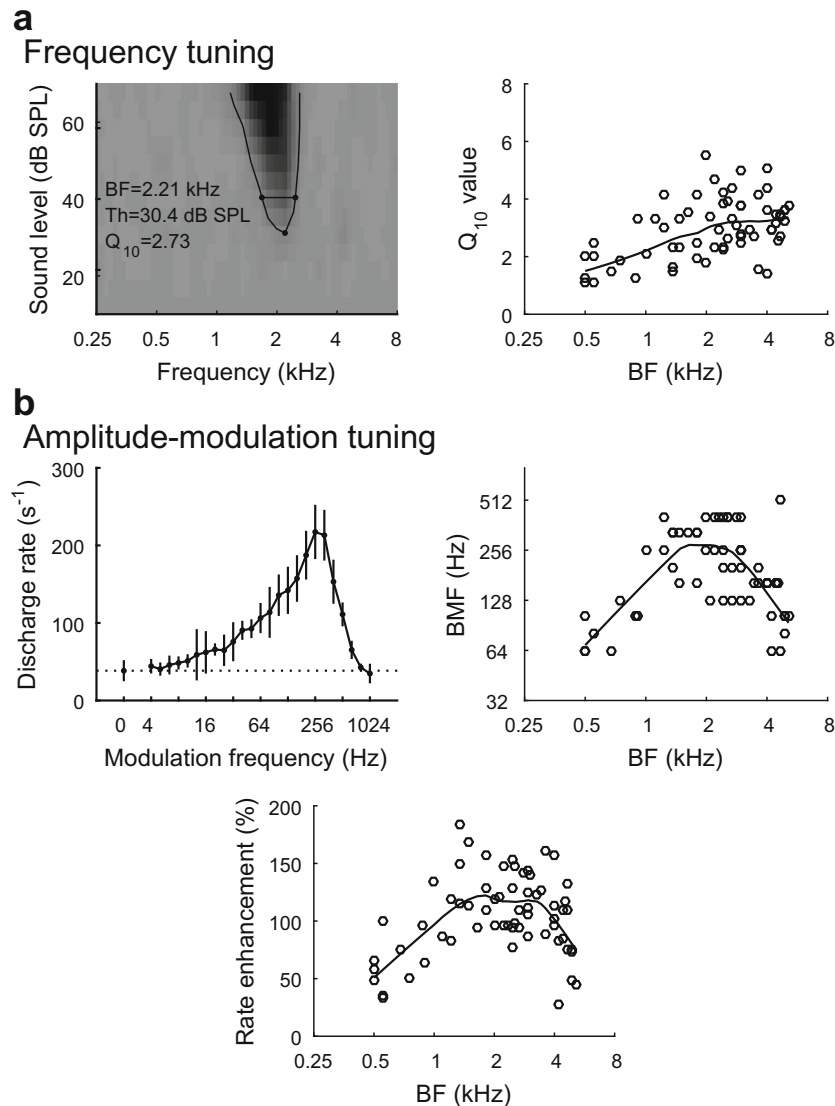


FIG. 3. Basic tuning properties of recording sites in the inferior colliculus (IC) of the budgerigar. **a** Representative frequency response map of one recording site (B45–10; *left*) showing average response rate to pure tones as a function of stimulus frequency and sound level. Darker shades of gray indicate higher response rate. Best frequency (BF), threshold (Th; in dB SPL), and Q_{10} value (BF/10-dB bandwidth), measured from the tuning curve, are given at the bottom *left*. The *right* panel shows the Q_{10} value of individual recording sites as a function of BF. **b** Representative modulation transfer function from the same recording site (B45–10; *top left*) showing average response rate (mean \pm SD) to AM tones as a function of modulation

frequency. AM tones were presented with 100 % modulation depth and carrier frequency of 2 kHz. The *dotted horizontal line* indicates the mean response rate to the unmodulated stimulus. Best modulation frequency (BMF) is 256 Hz. The *top right panel* shows BMF of individual recording sites as a function of BF. The *bottom panel* shows rate enhancement (percent difference between the maximum rate and the rate evoke by the unmodulated stimulus) of individual recording sites as a function of BF. Trend lines for 10-dB bandwidth, BMF, and rate enhancement were calculated using local regression analyses

to BF (1.8 kHz for the representative data shown in Fig. 4) but occurred instead when the formant frequency was approximately intermediate between two harmonics rather than aligned with a harmonic. The emergence of local rate peaks can be attributed to the pronounced F0-related envelope structure of between-harmonic stimuli (Fig. 1) combined with the band-enhanced modulation tuning properties of neurons in the IC (Fig. 3b).

Period histograms of IC responses to single-formant stimuli in quiet showed marked fluctuations in discharge rate over time associated with temporal synchrony to F0-related envelope structure (Fig. 4a). Response synchrony to individual frequency components was not observed. The amplitude of response synchrony to F0 was quantified through calculation of synchronized rate. Similar to average rate, synchronized rate to F0 exhibited local peaks when the

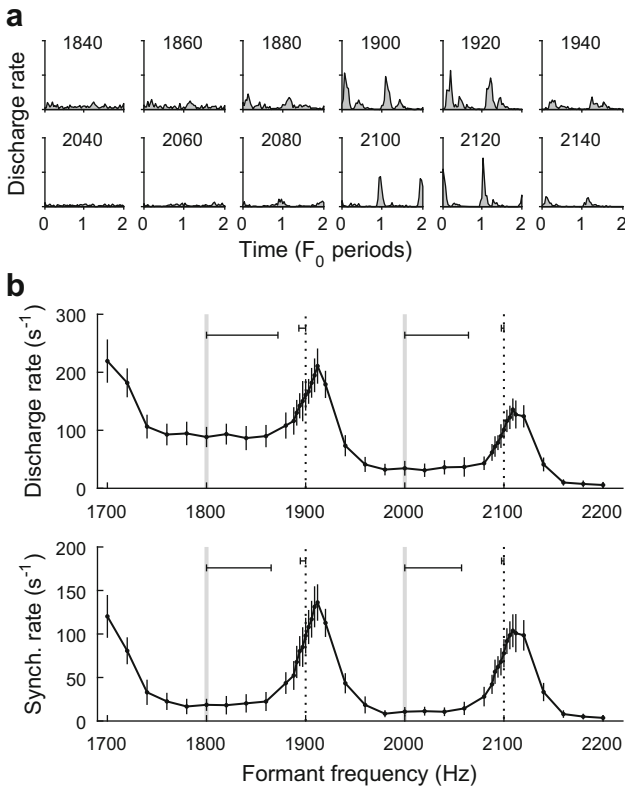


FIG. 4. Representative neural responses of a single recording site (B43–73; BF = 1.8 kHz; BMF = 323 Hz) to single-formant stimuli in quiet. **a** Period histograms showing temporal fluctuations in instantaneous discharge rate over two periods of F_0 . The frequency of the formant is indicated above each histogram. Y-axes extend from 0 to 2000 discharges per second. **b** Average rate (*top*) and synchronized rate to F_0 (*bottom*; means \pm SD) are plotted as a function of formant frequency. On- and between-harmonic frequencies are indicated with vertical gray lines and vertical dashed lines, respectively. Horizontal error bars show formant discrimination thresholds, based on ROC analyses, for the on- and between-harmonic standard stimuli. Large peaks in average rate and synchronized rate occur when formant frequency is approximately between two harmonic frequencies

formant frequency was approximately intermediate between two harmonics rather than aligned with a harmonic (Fig. 4b, bottom), consistent with the large envelope fluctuations of between-harmonic stimuli. Also similar to average discharge rate, synchronized rate increased gradually as the formant frequency of the stimulus approached BF.

Thresholds of individual neural recording sites for formant frequency discrimination in quiet were estimated based on average rate and synchronized rate to F_0 using ROC analysis. Neural thresholds for discrimination of behavior-matched stimuli were lower (more sensitive) when the peak frequency of the standard stimulus was between two harmonics rather than aligned with a harmonic (Fig. 5a; $N = 64$), consistent with the behavioral results. Neural thresholds were lower for the between-harmonic standard based on

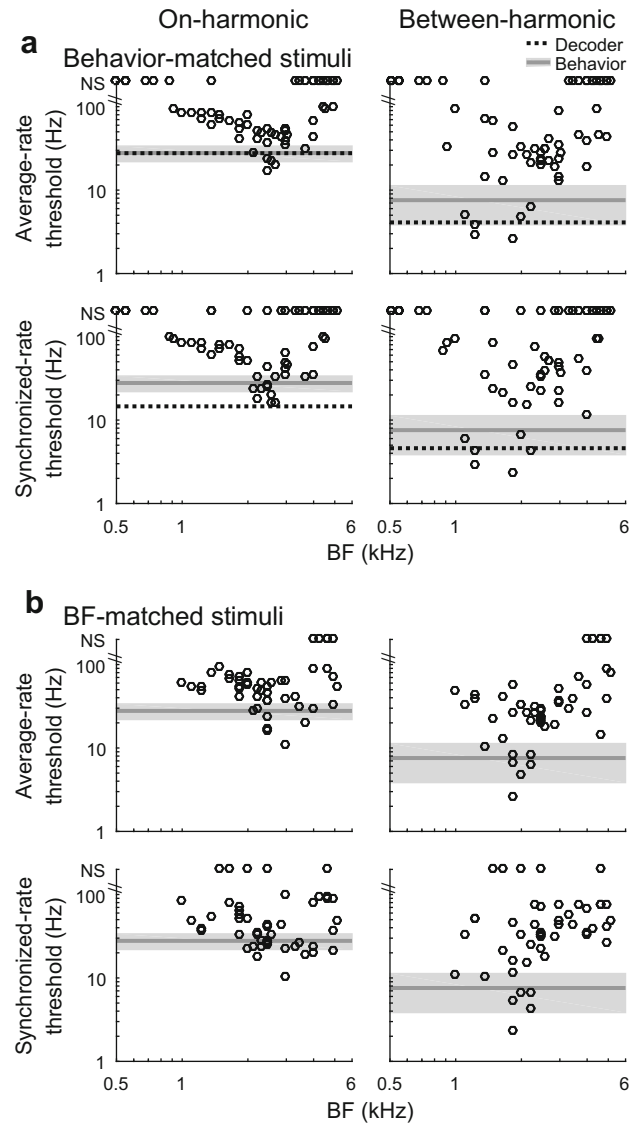


FIG. 5. Neural thresholds for discrimination of formant frequency in quiet. Thresholds are shown for the reference stimuli matched to behavioral experiments (**a**) and matched to the BF of the recording site (**b**). Neural thresholds of individual recording sites (*circles*) were calculated based on ROC analysis of average rate (*top*) and synchronized rate to F_0 (*bottom*) and are plotted as a function of BF along with thresholds of the optimally pooled neural population (*horizontal dashed lines*) and behavioral thresholds of the budgerigar (*gray*; means \pm SD). Thresholds falling along the tops of each plot are not significant ('NS'), i.e., exceed the upper limit of the estimation procedure

both average rate (Fig. 5a, top) and synchronized rate to F_0 (Fig. 5a, bottom), with similar thresholds observed between response metrics. The best thresholds within the neural population were sensitive enough to account for behavioral formant frequency discrimination abilities in the budgerigar and in humans. The most sensitive neural thresholds were observed at recording sites with BFs from 1 to 2 kHz for the between-harmonic standard stimulus and from 2 to 3 kHz for the on-harmonic

standard. Neural thresholds for BF-matched stimuli in quiet (Fig. 5b; $N = 48$) were similar to thresholds for behavior-matched stimuli.

The discrimination performance of the combined neural population was investigated for behavior-matched stimuli in quiet using a maximum likelihood-based pattern decoder (Jazayeri and Movshon 2006; Day and Delgutte 2013). These analyses optimally pool average rate or synchronized rate information across neural recording sites ($N = 64$). Thresholds of the pooled neural population for formant discrimination in quiet were approximately as low as the most sensitive individual recording sites in the population. Neural thresholds in quiet were lower for between-harmonic than on-harmonic standard stimuli, and neural thresholds based on average rate were similar to those based on synchronized rate to F₀ (Fig. 5a). Thresholds of the pooled neural population based on both response

metrics were sensitive enough to account for behavioral performance levels under quiet conditions.

Synchrony-Based Neural Thresholds Account for Behavioral Formant Discrimination Thresholds in Noise

For a subsample of recording sites studied in quiet, neural responses to behavior- and BF-matched single-formant stimuli were also recorded in simultaneously gated speech spectrum-shaped noise. In contrast to previous observations of responses to stimuli without background noise, changes in average discharge rate with formant frequency were almost invariably abolished in noise (Fig. 6b, top). Period histograms of IC responses in noise often showed synchrony to the F₀-related envelope structure of single-formant stimuli (Fig. 6a). Response envelope synchrony

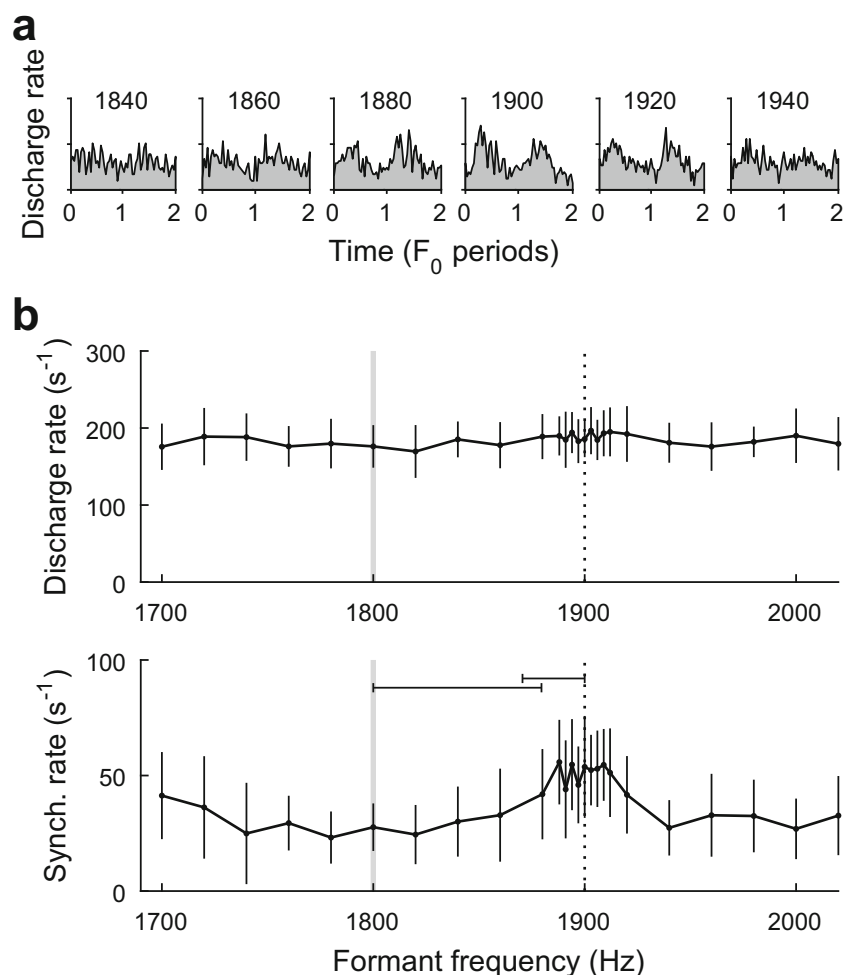


FIG. 6. Representative neural responses of a single recording site (B45-10; BF = 2.2 kHz; BMF = 256 Hz) to single-formant stimuli in noise. Period histograms (a) and average/synchronized rate to F₀ plotted as a function of formant frequency (b), as in Fig. 4. Y-axes of period histograms extend from 0 to 600 discharges per second.

Formant frequency discrimination thresholds based on average rate exceed the upper limit of the analysis procedure. Peaks in average discharge rate observed in quiet are abolished in noise. Peaks in synchronized rate to F₀ in quiet are dampened in noise

tended to peak when the frequency of the formant was between two harmonics rather than aligned with a harmonic (Fig. 6b, bottom), but local peaks in synchrony were dampened considerably compared to IC response patterns in quiet.

Average rate-based thresholds of individual recording sites for formant-frequency discrimination of behavior-matched stimuli in noise invariably exceeded the 100-Hz upper limit of the threshold estimation procedure (Fig. 7a, top; $N = 21$). Thresholds for BF-matched stimuli also exceeded 100 Hz, except for at one recording site with BF well above the spectral peak of the speech spectrum-shaped noise (BF = 4.9 kHz; Fig. 7b, top). In contrast to rate-based thresholds, thresholds based on synchronized rate to F0 could often still be measured in noise for BF-matched stimuli (Fig. 7b, bottom panels) and for behavior-matched stimuli when the BF of the recording site was near 2.0 kHz (Fig. 7a, bottom panels). The best synchrony-based neural thresholds of individual recording sites were sensitive enough to account for behavioral formant

discrimination abilities in noise. Synchrony-based thresholds were similar for on-harmonic and between-harmonic reference stimuli, in contrast with behavioral results.

Similar to the thresholds of individual recording sites, average-rate-based thresholds of the pooled neural population in noise also exceeded the upper limit of the estimation procedure ($N = 21$). In contrast, pooled thresholds based on synchronized rate to F0 were more sensitive (note that the threshold of the population is largely determined by the performance of three individual recording sites with significant individual thresholds and BFs similar to the stimulus test frequency). Pooled synchrony thresholds were lower for between-harmonic than for on-harmonic reference stimuli, consistent with behavioral data. The pooled synchrony threshold of the neural population was low enough to account for the behavioral discrimination threshold for the on-harmonic reference stimulus and approached the behavioral threshold for the between-harmonic reference stimulus in background noise (Fig. 7a).

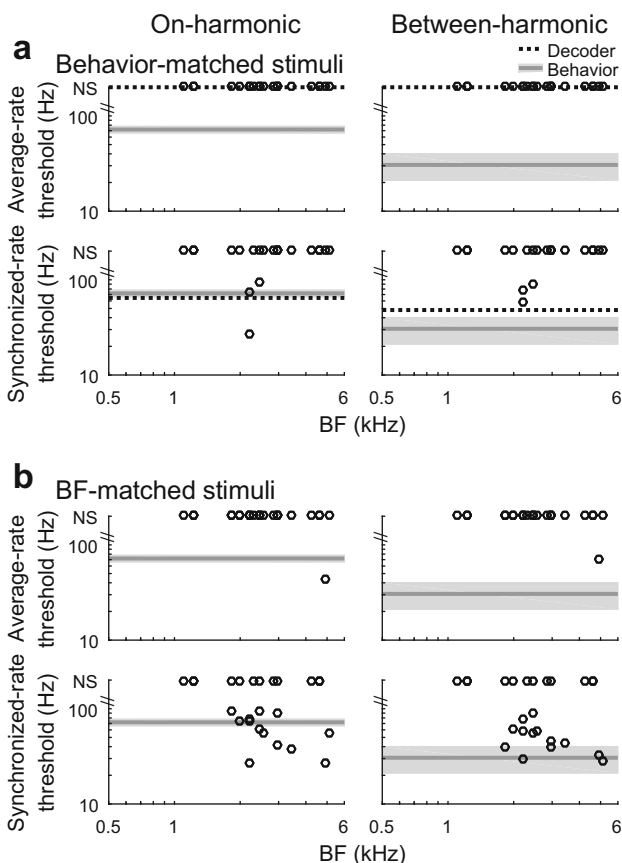


FIG. 7. Neural thresholds for discrimination of formant frequency in noise. Neural thresholds of individual recording sites and the optimally pooled population are shown, together with behavioral thresholds of the budgerigar as in Fig. 5, for behavioral (a) and BF-matched (b) reference stimuli

Rate-Coding of Formant Frequency Arises from Band-Enhanced Modulation Tuning

Phenomenological models of the auditory nerve (Zilany et al. 2014) and band-enhanced modulation tuning in the IC (Mao and Carney 2015) were used to gain insight into the contribution of rate-based modulation tuning to neural coding of formant frequency. Model auditory nerve fibers with characteristic frequencies near 2 kHz showed subtle fluctuations in average discharge rate with increasing formant frequency in quiet (Fig. 8, top left). In contrast, larger changes were observed in synchronized rate to F0-related envelope structure, which exhibited local peaks when formant frequency was approximately between harmonic frequencies rather than aligned with a harmonic (Fig. 8, bottom left).

Model IC neurons with BFs near 2 kHz showed substantial, non-monotonic changes in both average discharge rate and synchronized rate to F0 with increasing formant frequency in quiet that were consistent with observed response patterns in the IC (Fig. 8, right). That is, local peaks in average rate and synchronized rate occurred when formant frequency was approximately between harmonics rather than aligned with a harmonic. These results show that a sensitive, rate-based neural representation of formant frequency can emerge as a consequence of a simple filtering process that produces band-enhanced modulation tuning.

The addition of simultaneously gated speech spectrum-shaped noise abolished the small fluctuations as a function of formant frequency in model auditory nerve average response rate, as observed previously in

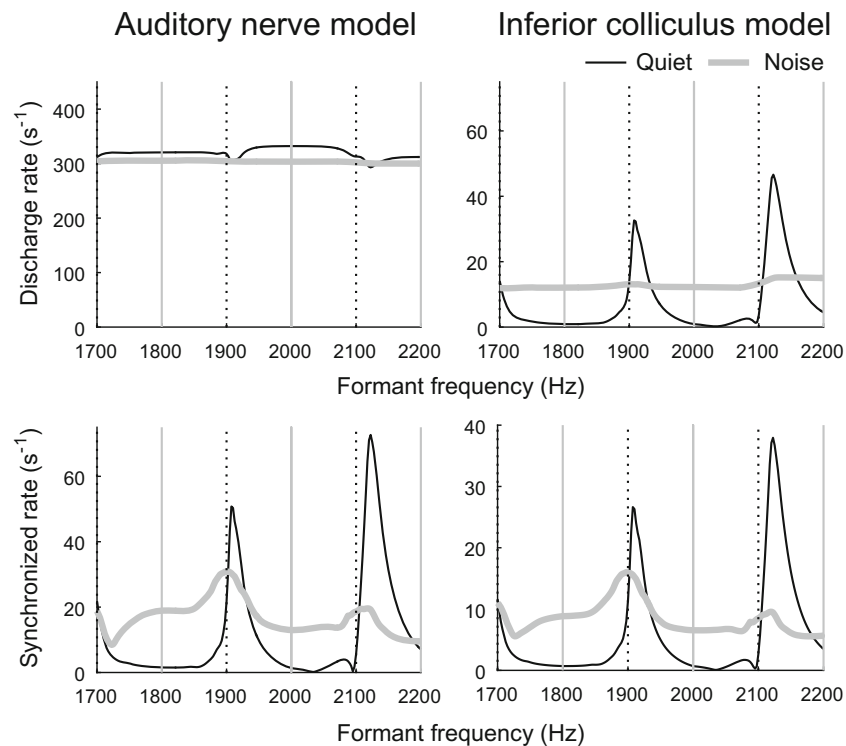


FIG. 8. Representative single-formant responses from models of the auditory nerve (*left*) and IC (*right*). Predicted average rate (*top*) and synchronized rate to F0 (*bottom*) at each model stage are plotted as a function of formant frequency for stimuli in quiet and in noise. Model BF was 1.8 kHz; BMF was 200 Hz. The auditory nerve stage predicts peaks in synchronized rate to F0, but not average rate, when

formant frequency is approximately between harmonics. IC model predictions are consistent with physiological data from the budgerigar: peaks in average rate occur approximately between harmonics in quiet but not in noise, while peaks in synchronized rate occur between harmonics both in quiet and in noise

quiet (Fig. 8, top left; gray traces). Fluctuations in auditory nerve synchronized rate to F0 in noise were reduced in noise compared to in quiet. Responses of model IC neurons to stimuli in background noise (Fig. 8, right; gray traces) showed minimal variation in average rate as a function of formant frequency but retained modest fluctuations in F0-related envelope synchrony. Response patterns of model IC neurons in noise were generally consistent with observed neurophysiological data.

DISCUSSION

The present study quantified neural and behavioral thresholds for discrimination of simplified vowel-like sounds in the budgerigar to gain insight into the coding mechanisms in the IC that support behavioral sensitivity to small changes in formant frequency. Behavioral formant frequency discrimination thresholds in budgerigars decreased with increasing amplitude of F0-related envelope cues, increased in noise, and were similar to human thresholds measured using matched procedures. Neurophysiological recordings in awake birds revealed that the IC encodes vowel-like sounds in quiet through both neural synchrony to F0-related envelope

structure and variation in average discharge rate (also related to envelope structure). However, whereas IC thresholds based on average rate were sensitive enough to support behavioral thresholds in quiet, only synchrony-based neural thresholds could account for behavioral thresholds both in quiet and in background noise. Finally, model simulations showed that average-rate coding of vowel-like sounds in quiet arises due to amplitude modulation tuning in the IC. Tuning of midbrain neurons to amplitude-modulation frequency is shared between birds and mammals.

IC thresholds based on average rate were sensitive enough to explain behavioral discrimination of single-formant stimuli in quiet, but not in noise. Neural discrimination thresholds in noise could not be estimated based on average discharge rates of either individual recording sites or of the optimally pooled neural population. In contrast, IC thresholds based on synchrony to envelope structure could explain behavioral discrimination thresholds both in quiet and in noise. These findings were further supported by modeling results, which predicted variation in synchronized rate to F0, but not average rate, across stimuli presented in noise. The results show that at the level of the IC, the information necessary for frequency discrimination of single-formant stimuli in

noise is encoded in the temporal discharge pattern of neural responses rather than through average discharge rate. Neural synchrony to envelope structure in the IC is also thought to support behavioral detection of AM tone- and noise-carrier stimuli. While rate-based AM detection thresholds are not sensitive enough to account for behavioral thresholds for modulation frequencies below 100 Hz, synchrony-based thresholds are sensitive enough to support behavioral thresholds across a broad range of modulation frequencies for both tone- and noise-carrier stimuli (Henry et al. 2016). Taken together, these studies underscore the behavioral significance of neural synchrony to envelope structure at the level of the IC. Whereas the previous AM study demonstrated the importance of envelope synchrony for detection of low-modulation frequencies, the present study highlights a role of envelope synchrony for formant frequency discrimination in background noise (where discrimination from the on-harmonic standard requires detection of an increase in modulation depth and discrimination from the between-harmonic standard requires detection of a decrease in modulation depth; see Fig. 1). The average-rate representation of single-formant stimuli in quiet was primarily a consequence of band-enhanced modulation tuning in the IC. Whereas more peripheral auditory nuclei encode envelope structure largely through response synchrony alone (Joris and Yin, 1992; Rhode and Greenberg, 1994; Gleich and Klump, 1995; Sayles et al. 2013), many IC neurons show band-enhanced modulation tuning in which average discharge rate increases with modulation depth for stimuli presented within a limited band of modulation frequencies (Langner and Schreiner 1988; Krishna and Semple 2000; Joris et al. 2004; Woolley and Casseday 2005; Nelson and Carney 2007; Henry et al. 2016). This basic response property explains why stimuli with the formant frequency centered approximately between harmonics evoked robust average rate responses from the IC: these sounds contain two dominant frequency components of equal amplitude that beat to produce large envelope fluctuations. Concomitantly, on-harmonic stimuli evoked lower rates because they are dominated by a single frequency component and hence contain relatively small envelope fluctuations. On a broader scale, average discharge rate in the IC also varied with the proximity of the formant frequency to BF. These findings show that response patterns in the IC cannot be understood based on either spectral tuning or modulation tuning alone but depend instead on the interaction of these tuning properties.

The contribution of band-enhanced modulation tuning in the IC to average-rate coding of vowel-like sounds was further supported by the modeling

results. Phenomenological models of the mammalian auditory nerve and IC (Zilany et al. 2014; Mao and Carney 2015) were able to reproduce the average-rate representation of formant frequency observed in our IC recordings, without any adjustment of model parameters, from a primarily synchrony-based representation in the peripheral input stage. Band-enhanced modulation tuning in this implementation of the IC model arises from simple band-pass filtering of the time-varying instantaneous discharge rate of a model auditory nerve fiber, but equivalently, could arise through putative fast excitation coupled with stronger, delayed inhibition from lemniscal input fibers with similar BFs (Nelson and Carney 2004; Nelson and Carney 2007; Mao and Carney 2015). This basic model structure appears physiologically plausible, considering that it reproduces commonly observed changes in AM response properties observed with pharmacological blockage of inhibitory inputs in the IC, including elevation of discharge rate across modulation frequencies with minimal change in BMF (Burger and Pollak 1998; Caspary et al. 2002; Zhang and Kelly 2003).

The modeling results also help explain why peaks in IC response amplitude were often observed *approximately* between harmonic frequencies rather than at exact midpoints. As previously noted, the deep envelope fluctuations of the 2100-Hz stimulus waveform are caused by beating between roughly equal-amplitude harmonics at 2000 and 2200 Hz (the components differ in amplitude by 0.9 dB for the 2100-Hz stimulus; equal amplitude is observed for the 2097.6-Hz stimulus). When these harmonics are sufficiently above BF (e.g., as in Fig. 4), cochlear filtering reduces the amplitude of the 2200-Hz harmonic more than of the 2000-Hz harmonic. This change reduces the amplitude of envelope fluctuations at the output of the cochlear filter and ultimately reduces response amplitude to the 2100 Hz stimulus in the IC. Response amplitude in the IC peaks instead at a slightly higher formant frequency, where deep envelope fluctuations are restored through a relative boost in the amplitude of the 2200-Hz component.

The average-rate representation of formant frequency observed in the budgerigar IC under quiet conditions probably also exists in mammals, considering that the models used here are based on neurophysiological data from mammals. The ability of these models to predict IC responses in the budgerigar highlights an emerging pattern of broad similarities in auditory function between birds and mammals up to at least the level of the auditory midbrain (Ryugo and Parks 2003; Woolley and Portfors 2013). While birds and mammals exhibit differences in the anatomy of the cochlea related to extension of the upper frequency limit of hearing in

mammals (Manley 2010), auditory nerve fibers in both groups show similar ranges of spontaneous activity, minimum thresholds for tone stimuli, frequency tuning bandwidth as a function of BF, and dynamic range/rate saturation (Sachs et al. 1974; Manley et al. 1985). Furthermore, auditory nerve fibers in birds, as in mammals, encode AM through envelope synchrony and not through changes in average discharge rate with modulation depth (Gleich and Klump 1995).

Band-enhanced modulation tuning is a common response property in the IC of both mammals and birds. The proportion of neurons with band-enhanced modulation tuning varies across species from approximately 90 % in the budgerigar (Henry et al. 2016) to 60–70 % in cat (Langner and Schreiner 1988) and 45–50 % in guinea pig and rabbit (Rees and Palmer 1989; Nelson and Carney 2007). The distribution of BMFs also varies across species, being somewhat higher in the budgerigar and chinchilla (typically 100–300 Hz; Langner et al. 2002; Henry et al. 2016) than in gerbil, rabbit, and cat (typically 30–100 Hz; Langner and Schreiner, 1988; Krishna and Semple, 2000; Nelson and Carney, 2007). However, the extent to which these patterns reflect species differences versus differences in methodology is unclear. Multiunit recordings, as used here and by others (Langner and Schreiner 1988; Langner et al. 2002), have been suggested to produce an upward shift in the distribution of BMFs, either because neurons with high BMFs are more difficult to isolate as single units or because these recordings possibly contain contributions from input fibers to the IC. Finally, the use of anesthesia (e.g., Langner and Schreiner, 1988; Krishna and Semple, 2000) could also influence modulation tuning in the IC due to effects on inhibition (Krishna and Semple 2000).

In summary, the present study demonstrated similar behavioral sensitivity to simplified vowel-like sounds in budgerigars and humans. While the average-rate representation of single-formant stimuli in the budgerigar IC was sensitive enough to support behavioral discrimination thresholds under quiet conditions, only the representation based on envelope synchrony could explain behavioral thresholds both in quiet and in noise. These findings indicate an incomplete transition to average-rate coding of vowel-like sounds in the IC and underscore the significance of temporal discharge patterns for perception of complex sounds. Furthermore, they highlight the potential benefits of incorporating F0-related envelope structure into stimulation strategies for auditory midbrain implants, which currently provide inadequate temporal information for robust speech perception (Lim and Lenarz 2015). Finally, model simulations showed that average-rate coding of

vowel-like sounds in quiet emerges from a synchrony-based representation in more peripheral nuclei due to amplitude-modulation tuning. This transformation likely also occurs in the mammalian IC and may lay the groundwork for reported average-rate coding of naturally spoken vowels in regions of the auditory cortex (Mesgarani et al. 2008).

ACKNOWLEDGMENTS

This work was supported by National Institutes of Health Grants R01-DC001641 to L.H.C. and R00-DC013792 to K.S.H. Mitchell L. Day provided analysis code for calculating the discrimination performance of optimally pooled neural populations.

Author Contributions KSH, JF, FI, and LHC designed the research; KSH, KSA, MJM, and EGN performed the experiments; KSH, KSA, MJM, FI, and LHC analyzed the data; and KSH wrote the manuscript.

COMPLIANCE WITH ETHICAL STANDARDS

Conflict of Interest The authors declare that they have no conflict of interest.

REFERENCES

- AITKIN LM, PHILLIPS SC (1984) Is the inferior colliculus an obligatory relay in the cat auditory system? *Neurosci Lett* 44:259–264
- BARTLETT EL, WANG X (2007) Neural representations of temporally modulated signals in the auditory thalamus of awake primates. *J Neurophysiol* 97:1005–1017. doi:10.1152/jn.00593.2006
- BAUMANN S, GRIFFITHS TD, SUN L, PETKOV CI, THIELE A, REES A (2011) Orthogonal representation of sound dimensions in the primate midbrain. *Nat Neurosci* 14:423–425. doi:10.1038/nn.2771
- BURGER RM, POLLAK GD (1998) Analysis of the role of inhibition in shaping responses to sinusoidally amplitude-modulated signals in the inferior colliculus. *J Neurophysiol* 80:1686–1701
- BYRNE D (1994) An international comparison of long-term average speech spectra. *J Acoust Soc Am* 96:2108. doi:10.1121/1.410152
- CALFORD MB, WISE LZ, PETTIGREW JD (1985) Coding of sound location and frequency in the auditory midbrain of diurnal birds of prey, families accipitridae and falconidae. *J Comp Physiol A* 157:149–160. doi:10.1007/BF01350024
- CARNEY LH, KETTERER AD, ABRAMS KS, SCHWARZ DM, IDROBO F (2013) Detection thresholds for amplitude modulations of tones in budgerigar, rabbit, and human. *Adv Exp Med Biol* 787:391–398. doi:10.1007/978-1-4614-1590-9_43
- CARNEY LH, LI T, McDONOUGH JM (2015) Speech coding in the brain: representation of vowel formants by midbrain neurons tuned to sound fluctuations. *ENEURO* 2:1–12. doi:10.1523/ENEURO.0004-15.2015
- CASPARY DM, PALOMBI PS, HUGHES LF (2002) GABAergic inputs shape responses to amplitude modulated stimuli in the inferior colliculus. *Hear Res* 168:163–173

- CHEN C, RODRIGUEZ FC, READ HL, ESCABI MA (2012) Spectrotemporal sound preferences of neighboring inferior colliculus neurons: implications for local circuitry and processing. *Front Neural Circuits* 6:62. doi:10.3389/fncir.2012.00062
- CHOI JH, JUNG HK, KIM T (2006) A new action potential detector using the MTEO and its effects on spike sorting systems at low signal-to-noise ratios. *IEEE Trans Biomed Eng* 53:738–746. doi:10.1109/TBME.2006.870239
- DAY ML, DELGUTTE B (2013) Decoding sound source location and separation using neural population activity patterns. *J Neurosci* 33:15837–15847. doi:10.1523/jneurosci.2034-13.2013
- DELGUTTE B, KIANG YS (1984) Speech coding in the auditory nerve: I. Vowel-like sounds. *J Acoust Soc Am* 75:866–878
- DOOLING RJ, BROWN SD (1990) Speech perception by budgerigars (*Melopsittacus undulatus*): spoken vowels. *Percept Psychophys* 47:568–574
- DOOLING RJ, SEARCY MH (1981) Amplitude modulation thresholds for the parakeet (*Melopsittacus undulatus*). *J Comp Physiol A* 143:383–388. doi:10.1007/BF00611177
- DOOLING RJ, OKANOVA K, BROWN SD (1989) Speech perception by budgerigars (*Melopsittacus undulatus*): the voiced-voiceless distinction. *Percept Psychophys* 46:65–71
- DOOLING RJ, BEST CT, BROWN SD (1995) Discrimination of synthetic full-formant and sinewave/ra-la/continua by budgerigars (*Melopsittacus undulatus*) and zebra finches (*Taeniopygia guttata*). *J Acoust Soc Am* 97:1839–1846
- DOOLING RJ, LOHR B, DENT ML (2000) Hearing in birds and reptiles. In: Dooling RJ, Fay RR, Popper AN (eds) *Comparative hearing: birds and reptiles*. Springer, New York, pp. 308–359
- EGAN JP (1975) *Signal detection theory and ROC analysis*. Academic Press, New York
- FANT G (1960) *Acoustic theory of speech production*. Mouton, Hague, The Netherlands
- GEORGE ELJ, FESTEN JM, HOUTGAST T (2006) Factors affecting masking release for speech in modulated noise for normal-hearing and hearing-impaired listeners. *J Acoust Soc Am* 120:2295–2311. doi:10.1121/1.2266530
- GERSHON ED, WIENER MC, LATHAM PE, RICHMOND BJ (1998) Coding strategies in monkey VI and inferior temporal cortices. *J Neurophysiol* 79:1135–1144
- GLEICH O, KLUMP GM (1995) Temporal modulation transfer functions in the European starling (*Sturnus vulgaris*): II. Responses of auditory nerve fibers. *Hear Res* 82:81–92
- HENRY KS, NEILANS EG, ABRAMS KS, IDROBO F, CARNEY LH (2016) Neural correlates of behavioral amplitude modulation sensitivity in the budgerigar midbrain. *J Neurophysiol* 115:1905–1916. doi:10.1152/jn.01003.2015
- HILLENBRAND J, GETTY L A, CLARK MJ, WHEELER K (1995) Acoustic characteristics of American English vowels. *J Acoust Soc Am* 97:3099–3111. doi:10.1121/1.411872
- HONEY C, SCHNUPP J (2015) Neural resolution of formant frequencies in the primary auditory cortex of rats. *PLoS One* 10:1–20. doi:10.1371/journal.pone.0134078
- JAZAYERI M, MOVSHON JA (2006) Optimal representation of sensory information by neural populations. *Nat Neurosci* 9:690–696. doi:10.1038/nn1691
- JOHNSON JS, YIN P, O'CONNOR KN, SUTTER ML (2012) Ability of primary auditory cortical neurons to detect amplitude modulation with rate and temporal codes: neurometric analysis. *J Neurophysiol* 107:3325–3341. doi:10.1152/jn.00812.2011
- JORIS PX, YIN TCT (1992) Responses to amplitude-modulated tones in the auditory nerve of the cat. *J Acoust Soc Am* 91:215–232. doi:10.1121/1.402757
- JORIS PX, SCHREINER CE, REES A (2004) Neural processing of amplitude-modulated sounds. *Physiol Rev* 84:541–577. doi:10.1152/physrev.00029.2003
- KEWLEY-PORT D, WATSON CS (1994) Formant-frequency discrimination for isolated English vowels. *J Acoust Soc Am* 95:485–496. doi:10.1121/1.410024
- KEWLEY-PORT D, BURKLE TZ, LEE JH (2007) Contribution of consonant versus vowel information to sentence intelligibility for young normal-hearing and elderly hearing-impaired listeners. *J Acoust Soc Am* 122:2365–2375. doi:10.1121/1.2773986
- KIM KH, KIM SJ (2000) Neural spike sorting under nearly 0-dB signal-to-noise ratio using nonlinear energy operator and artificial neural-network classifier. *IEEE Trans Biomed Eng* 47:1406–1411. doi:10.1109/10.871415
- KLATT DH, KLATT LC (1990) Analysis, synthesis, and perception of voice quality variations among female and male talkers.
- KRISHNA BS, SEMPLE MN (2000) Auditory temporal processing: responses to sinusoidally amplitude-modulated tones in the inferior colliculus. *J Neurophysiol* 84:255–273
- LADEFOGED P, MADDISON I (1996) *The sounds of the World's languages*. Wiley-Blackwell, Hoboken, NJ
- LANGNER G, SCHREINER CE (1988) Periodicity coding in the inferior colliculus of the cat. I. Neuronal mechanisms. *J Neurophysiol* 60:1799–1822
- LANGNER G, ALBERT M, BRIEDE T (2002) Temporal and spatial coding of periodicity information in the inferior colliculus of awake chinchilla (*Chinchilla laniger*). *Hear Res* 168:110–130. doi:10.1016/S0378-5955(02)00367-2
- LEVITT H (1970) Transformed up-down methods in psychoacoustics. *J Acoust Soc Am* 49:467–477
- LIM HH, LENARZ T (2015) Auditory midbrain implant: research and development towards a second clinical trial. *Hear Res* 322:212–223. doi:10.1016/j.heares.2015.01.006
- LYZENGA J, HORST JW (1995) Frequency discrimination of bandlimited harmonic complexes related to vowel formants. *J Acoust Soc Am* 98:1943–1955
- LYZENGA J, HORST JW (1997) Frequency discrimination of stylized synthetic vowels with a single formant. *J Acoust Soc Am* 102:1755–1767
- LYZENGA J, HORST JW (1998) Frequency discrimination of stylized synthetic vowels with two formants. *J Acoust Soc Am* 104:2956–2966. doi:10.1121/1.423878
- MACMILLAN NA, CREELMAN CD (2005) *Detection theory: a user's guide*.
- MANLEY GA (2010) An evolutionary perspective on middle ears. *Hear Res* 263:3–8. doi:10.1016/j.heares.2009.09.004
- MANLEY GA, GLEICH O, LEPPELSACK HJ, OECKINGHAUS H (1985) Activity patterns of cochlear ganglion neurones in the starling. *J Comp Physiol A* 157:161–181. doi:10.1007/BF01350025
- MAO J, CARNEY LH (2015) Tone-in-noise detection using envelope cues: comparison of signal-processing-based and physiological models. *J Assoc Res Otolaryngol* 16:121–133. doi:10.1007/s10162-014-0489-1
- MESGARANI N, DAVID SV, FRITZ JB, SHAMMA SA (2008) Phoneme representation and classification in primary auditory cortex. *J Acoust Soc Am* 123:899. doi:10.1121/1.2816572
- NELSON PC, CARNEY LH (2004) A phenomenological model of peripheral and central neural responses to amplitude-modulated tones. *J Acoust Soc Am* 116:2173–2186. doi:10.1121/1.1784442
- NELSON PC, CARNEY LH (2007) Neural rate and timing cues for detection and discrimination of amplitude-modulated tones in the awake rabbit inferior colliculus. *J Neurophysiol* 97:522–539. doi:10.1152/jn.00776.2006
- PEREZ CA, ENGINEER CT, JAKKAMSETTI V, CARRAWAY RS, PERRY MS, KILGARD MP (2013) Different timescales for the neural coding of consonant and vowel sounds. *Cereb Cortex* 23:670–683. doi:10.1093/cercor/bhs045

- RAMACHANDRAN R, DAVIS KA, MAY BJ (1999) Single-unit responses in the inferior colliculus of decerebrate cats. I. Classification based on frequency response maps. *J Neurophysiol* 82:152–163
- REES A, PALMER AR (1989) Neuronal responses to amplitude-modulated and pure-tone stimuli in the Guinea pig inferior colliculus, and their modification by broadband noise. *J Acoust Soc Am* 85:1978–1994
- RHODE WS, GREENBERG S (1994) Encoding of amplitude modulation in the cochlear nucleus of the cat. *J Neurophysiol* 71:1797–1825
- ROBBINS H, MONRO S (1951) A stochastic approximation method. *Ann Math Stat* 22:400–407
- ROSEN MJ, SEMPLE MN, SANES DH (2010) Exploiting development to evaluate auditory encoding of amplitude modulation. *J Neurosci* 30:15509–15520. doi:10.1523/JNEUROSCI.3340-10.2010
- RYUGO DK, PARKS TN (2003) Primary innervation of the avian and mammalian cochlear nucleus. *Brain Res Bull* 60:435–456. doi:10.1016/S0361-9230(03)00049-2
- SACHS MB, YOUNG ED (1980) Effects of nonlinearities on speech encoding in the auditory nerve. *J Acoust Soc Am* 68:858–875
- SACHS MB, YOUNG ED, LEWIS RH (1974) Discharge patterns of single fibers in the pigeon auditory nerve. *Brain Res* 70:431–447
- SACHS MB, VOIGT HF, YOUNG ED (1983) Auditory nerve representation of vowels in background noise. *J Neurophysiol* 50:27–45
- SAYLES M, FÜLLGRABE C, WINTER IM (2013) Neurometric amplitude-modulation detection threshold in the Guinea-pig ventral cochlear nucleus. *J Physiol* 591:3401–3419. doi:10.1113/jphysiol.2013.253062
- SESHAGIRI CV, DELGUTTE B (2007) Response properties of neighboring neurons in the auditory midbrain for pure-tone stimulation: a tetrode study. *J Neurophysiol* 98:2058–2073. doi:10.1152/jn.01317.2006
- TAN Q, CARNEY LH (2005) Encoding of vowel-like sounds in the auditory nerve: model predictions of discrimination performance. *J Acoust Soc Am* 117:1210–1222. doi:10.1121/1.1856391
- WOOLLEY SMN, CASSEDAY JH (2005) Processing of modulated sounds in the zebra finch auditory midbrain: responses to noise, frequency sweeps, and sinusoidal amplitude modulations. *J Neurophysiol* 94:1143–1157. doi:10.1152/jn.01064.2004
- WOOLLEY SMN, PORTFORS CV (2013) Conserved mechanisms of vocalization coding in mammalian and songbird auditory midbrain. *Hear Res* 305:45–56. doi:10.1016/j.heares.2013.05.005
- YIN P, JOHNSON JS, O'CONNOR KN, SUTTER ML (2011) Coding of amplitude modulation in primary auditory cortex. *J Neurophysiol* 105:582–600. doi:10.1152/jn.00621.2010
- YOUNG ED, SACHS MB (1979) Representation of steady-state vowels in the temporal aspects of the discharge patterns of populations of auditory-nerve fibers. *J Acoust Soc Am* 66:1381–1403
- ZHANG H, KELLY JB (2003) Glutamatergic and GABAergic regulation of neural responses in inferior colliculus to amplitude-modulated sounds. *J Neurophysiol* 90:477–490. doi:10.1152/jn.01084.2002
- ZILANY MSA, BRUCE IC, CARNEY LH (2014) Updated parameters and expanded simulation options for a model of the auditory periphery. *J Acoust Soc Am* 135:283–286. doi:10.1121/1.4837815