TECHNICAL RESEARCH REPORT

Perception and neural coding of harmonic fusion in ferrets

by Sridhar Kalluri, Shihab Shamma

CAAR TR 2004-1
(ISR TR 2004-23)

**Center for Auditory
and Acoustic Research**

Perception and cortical neural coding of harmonic fusion in ferrets

Sridhar Kalluri*

Institute for Systems Research

University of Maryland

College Park, MD 20742

(301) 405-7485, FAX: (301) 314-9920

sridhark@umd.edu

Shihab A. Shamma

Institute for Systems Research

Department of Electrical Engineering

University of Maryland

(301) 405-6842, FAX: (301) 314-9920

sas@glue.umd.edu

May 17, 2004

RUNNING TITLE: Perception and neural coding of harmonicity

*Corresponding author

I

The cortical neural correlates for the perception of harmonic sounds have remained a puzzle despite intense study over several decades. This study approached the problem from the point of view of the spectral fusion evoked by such sounds. Experiment 1 tested whether ferrets automatically fuse harmonic complex tones. In baseline sessions, three ferrets were trained to detect a pure tone terminating a sequence of inharmonic complex tones. After the ferrets reached proficiency in the baseline task, a small fraction of the inharmonic complex tones were replaced with harmonic tones. Two out of three ferrets confused the harmonic complex tones with the pure tones and responded as if detecting the pure tone at twice the false-alarm rate, indicating that ferrets can automatically fuse the partials of a harmonic complex. Experiment 2 sought correlates of harmonic fusion in single units of ferret primary auditory cortex (AI), by contrasting responses to harmonic complex tones with those to inharmonic complex tones. The effects of spectrotemporal filtering were accounted for by using the measured spectrotemporal receptive field to predict responses and by seeking correlates of harmonic fusion in the predictability of the responses. Ten percent of units exhibited some correlates of harmonic fusion, which is consistent with previous findings that no special processing for harmonic stimuli occurs in AI.

I.   Introduction

Two fundamentally important auditory perceptual phenomena, spectral fusion and periodicity pitch, are intimately associated with sounds having harmonic spectra. The importance of harmonic sounds in auditory perception is such that auditory theory for at least one hundred and fifty years has been driven in part by a quest for understanding the mechanisms underlying pitch. Despite intense investigation, many aspects of pitch perception have resisted explanation. One such problem is identifying its cortical neural correlates. More generally, not just pitch but the cortical encoding of harmonic sounds is not well understood, and forms the topic of this paper.

Listeners typically hear a harmonic complex tone as a coherent unitary entity with a clear pitch; this perceptual fusion due to harmonicity is used by the brain to organize complex acoustic environments into different auditory objects (Bregman, 1990). An example is the improvement in the ability to distinguish two talkers when a fundamental frequency difference is imposed (Brokx and Nooteboom, 1980). The highly salient pitch of harmonic complex tones is known as periodicity pitch. It is the most important of the many distinct percepts that come under the rubric of pitch, because periodicity pitch underlies speakers' voices and speech prosody, as well as musical intervals and melody.

The percepts of periodicity pitch and the spectral fusion of complex tones are closely related. There are a great many perceptual parallels, as in their similar insensitivity to the phases of low harmonics (Moore and Glasberg, 1985; Houtsma and Smurzynski, 1990; Hartmann, McAdams, and Smith, 1993). Moreover, models for periodicity pitch apply well to harmonic spectral fusion as well. For instance, in pattern recognition models, the pitch of a complex tone is equal to the fundamental frequency of the harmonic template that best fits the evoked neural activity arrayed along the tonotopic axis of the cochlea. In such models, the template-matching operation works on the neural activity due to the resolved partials of the complex tone (e.g., Wightman, 1973; Goldstein, 1973; Terhardt, 1974; Terhardt, 1979). There is evidence that a template-matching operation also underlies the perceptual fusion of harmonically related partials of a complex tone (Brunstrom and Roberts, 1998; Lin and Hartmann, 1998; Brunstrom and Roberts, 2000).

While spectral fusion and periodicity pitch of complex tones are closely related, previous stud-

ies of the encoding of harmonic sounds have focused on the latter aspect of the perception of complex tones, especially in auditory cortex. Lesion studies show that the auditory cortex is needed for the perception of periodicity pitch (Whitfield, 1980; Zatorre, 1988). However, studies of primary auditory cortex (AI) have failed, on the whole, to yield compelling correlates of periodicity pitch. Recordings from single units of monkey AI in response to harmonic complex tones with and without the missing fundamental component failed to show any rate tuning to the fundamental frequency (Schwarz and Tomlinson, 1990); such tuning to the fundamental frequency regardless of the physical presence of that component would have indicated that AI neurons are tuned to the pitch of harmonic complex tones. Multi-unit activity and current-source-density patterns recorded in high-frequency areas of monkey AI directly encode the click rate of same- and alternating-polarity click trains in the temporal pattern of response (Steinschneider et al., 1994); this click rate corresponds to the residue pitch, which is a weak pitch due to the unresolved harmonics of the stimulus. Consistent with studies of single-unit activity (Schwarz and Tomlinson, 1990), resolved harmonics were represented as local maxima of activity determined by the tonotopic organization of the recording sites. However, the periodicity pitch that would be derived from these local maxima was not encoded directly, in that neither the temporal pattern of response nor spatial distribution of activity reflected the fundamental frequency. Finally, several studies have reported mapping of the envelope periodicity of amplitude-modulated tones on an axis orthogonal to the tonotopic axis in AI of Mongolian gerbils (Schulze and Langner, 1997a; Schulze and Langner, 1997b; Schulze et al., 2002) and of humans (Langner et al., 1997). These findings may indicate the presence of a map of periodicity pitch, but two aspects of the experiments make the assertion inconclusive.

1. An extensive body of psychoacoustical literature shows that envelope periodicity, in general, is not predictive of the periodicity pitch evoked by a stimulus (e.g., de Boer, 1956; Flanagan and Gutman, 1960; de Boer, 1976).

2. Rather than reflecting the periodicity pitch estimated from the resolved components of the stimulus, the findings could instead be a mapping of the fundamental frequency (or modulation frequency for an amplitude-modulated tone) re-introduced by nonlinear distortion in

the cochlea (McAlpine, 2002).

Sensitivity to harmonic combinations of resolved partials that would show tuning to the fundamental frequency of complex tones have been found in marmoset AI (Kadia and Wang, 2003) as well as in bats. Such neurons may underlie spectral fusion of harmonic complex tones. However, because these neurons preferred high frequencies and very high fundamental frequencies (typically greater than 4 kHz) , the role of such neurons in the formation of periodicity pitch and spectral fusion is unclear for sounds having predominantly low-frequency spectra such as speech, music, and many animal vocalizations.

In summary, previous investigations of the encoding of harmonic sounds in AI have focused on correlates of periodicity pitch. It is reasonable to say that these studies have failed to find evidence that responses of AI neurons directly reflect the periodicity pitch of harmonic complex tones.

In this paper, we report experiments in domestic ferrets (*Mustela putorius*) aimed at understanding the cortical neural coding of harmonic sounds from the perspective of spectral fusion rather than periodicity pitch. By comparing the neural representation of perceptually fused harmonic complex tones with that of perceptually unfused inharmonic complex tones in single AI neurons, we expect to reveal neural computations specifically elicited by harmonic fusion in subcortical or primary cortical structures. The presence of such harmonicity-specific processing would have been missed by previous studies that employed harmonic sounds exclusively.

A basic assumption underlying the neurophysiology experiment is that ferrets automatically fuse partials of harmonic complex tones, like humans do in typical listening conditions. Many animals can hear the pitch of the missing fundamental, including cats and monkeys (Heffner and Whitfield, 1976; Tomlinson and Schwartz, 1988), so it is not unreasonable to assume that ferrets might do the same. In order to hear the pitch of the missing fundamental, the primate and feline subjects must have been able to estimate the fundamental from the components that were actually present in the stimuli. However, the animals did not necessarily fuse these components into a unitary entity[1]. The first experiment in this paper uses a behavioral-testing paradigm to specifically

---

[1]Spectral fusion of partials and their contribution to periodicity pitch are not entirely congruent in human listeners as well(Brunstrom and Roberts, 1998). For example, a partial must be mistuned from a harmonic relation by 1.5% to

test the assumption that ferrets automatically fuse harmonic partials of complex tones. Experiment 2 seeks neural correlates of harmonic fusion in ferret AI.

II. Experiment 1: Perception of harmonic complex tones

A. Rationale

In order to determine if ferrets can automatically fuse partials of a harmonic complex, we asked if they can distinguish inharmonic complex tones from harmonic complex tones without receiving training on the distinction between the two classes of sounds. The experiment was performed in two stages comprising baseline sessions and probe sessions (Figure 1). In baseline sessions, the ferrets were trained to detect pure-tone targets in a sequence of inharmonic-complex-tone reference sounds. By eliminating or making unreliable the differences in frequency ranges, levels, and roughnesses, two cues were left available to reliably distinguish targets from references:

1. Differences in the degree of perceptual fusion;

2. Differences in timbre.

When ferrets attained proficiency in these baseline sessions, probe sessions were conducted where 10% of the reference sounds were replaced with harmonic-complex-tone probe sounds that were identical to the inharmonic complex-tone references in every way except in the frequency relations between partials. Because the timbre of the complex tones differs greatly from that of the pure tones, ferrets were expected, in most cases, to categorize the harmonic complex tones with the inharmonic complex tones based on similarity of timbre. This should be the case even if the ferrets heard the harmonic complex tones as fused. However, occasionally the putative fused nature of the harmonic complex tone might be confused with the unitary quality of the pure tone, prompting the ferret to respond to the harmonic probes as if it heard the pure-tone target. Therefore, detection

---

2% in order to hear it apart from the remaining components, whereas it must be mistuned by 8% to stop contributing to the pitch of the complex (Moore and Kitzes, 1985; Moore, Peters, and Glasberg, 1986; Darwin and Ciocca, 1992; Ciocca and Darwin, 1993).

rate of the probes greater than the rate at which references were inadvertently detected indicates that harmonic complex tones are heard differently than inharmonic complex by eliciting a fused percept. In contrast, failure to detect the probes in this experiment does not mean that the ferrets could not fuse harmonic tones; they could simply be using timbre exclusively to categorize the stimuli.

[Insert Figure 1 Here]

B. Methods

1. **Psychoacoustical testing**

A conditioned-avoidance paradigm was used for testing how ferrets hear complex tones. The paradigm has been successfully used with many animals and it is described in detail by Heffner and Heffner (1995) . We give a brief overview here.

A water-deprived ferret licks water from a continuously dripping spout at one end of a testing cage while listening to reference sounds. At random intervals, an easily distinguishable target sound is presented followed by a light shock to the tongue delivered through the reward spout. Such pairings of stimulus and shock help the ferret learn to break contact with the spout when it hears a target. The continuous water reward encourages the ferret to maintain contact with the spout, providing a baseline behavior against which to measure responses. From the perspective of a ferret, the reference stimuli constitute safe trials because it can drink from the spout without getting shocked. On the other hand, target stimuli are warning trials, because they warn the ferret to break contact with the spout.

A computer registers successful withdrawal from the spout following a target as a *hit* and failure to withdraw as a *miss*. Because the animal occasionally withdraws from the spout in the absence of a target, a *false-alarm* rate is determined by registering how often it withdraws from the reward spout for reference trials. To ensure that only trials on which a ferret is attending to the stimulus are included for evaluating performance, both target and reference trials are ignored (and responses registered as *early* withdrawals) on which the ferret did not contact the reward spout immediately

before the trial. Figure 2 helps understand these measures by illustrating the presentation of trials and the timing of response intervals, while Table I shows which behaviors lead to the different response categories.

During testing, two kinds of trial sequences are presented.

1. One to six reference trials followed by a target trial;

2. Seven consecutive reference trials constituting a sham sequence.

The number of reference trials in a given sequence are randomized such that the probability of the target sound occurring in trial position 2 through 7 is constant, so that the ferret cannot preferentially respond on trials occurring at any given position. Sham sequences are interspersed between target sequences to discourage the ferrets from responding at regular intervals regardless of whether a target was presented. During probe sessions, probe trials are presented in exactly the same way as reference trials by replacing 10% of reference sounds by probe sounds. Responses on the probe trials are scored in the same way as those on the reference trials. *Probe hits* and *probe misses* are equivalent to false alarms and safe responses respectively on reference trials.

[Insert Figure 2 Here]

[Insert Table I Here]

2. **Stimuli**

For any given reference, target, or probe trial, stimuli were chosen randomly from a collection of inharmonic, pure-tone, and harmonic sounds that were synthesized and stored in computer memory just before placing a ferret in the testing cage. Pure tones ranged in frequency from 150 Hz to 4800 Hz. These frequencies were also the lower and upper bounds of the spectra of the complex tones, which had 6 components. To comply with these frequency limits, harmonic complex tones with components in random phase had fundamental frequencies between 150 Hz and 800 Hz. Inharmonic complex tones were synthesized by scrambling ratios between consecutive partials of harmonic complex tones so that the two types of complex tones elicited comparable percepts of roughness. The lowest partial of the inharmonic complex tones was constrained between 150 Hz

6

and 800 Hz, so that any pitch cues from the edge of the spectrum were not reliably different from those generated by the spectral edge of the harmonic complex tones. The levels of all stimuli were roved over a 6 dB range to ensure that intensity cues were not used. Stimuli were played at a nominal level of 70 dB SPL, calibrated in an empty testing cage at the position occupied by the ferret's head with a Bruel and Kjaer free-field microphone. All stimulus parameters were restricted to narrower ranges during training sessions to help the ferrets learn the task.

3. **Experimental apparatus**

Behavior of the ferrets was tested in a custom-designed cage mounted inside a Sonex-foam lined and single-walled sound-proof booth (Industrial Acoustics, Inc.). Sound was delivered through a speaker (Manger) mounted in the front of the cage at approximately the same height above the testing cage as the metal spout that delivered the water reward.

The testing cage had a metal floor so that the ferret formed a low-resistance electrical pathway between the spout and the floor when licking. The lowered resistance between floor and spout was used by a custom "touch" circuit to register when the ferret contacted the spout. Electro-mechanical relays switched between the touch circuit and a fence charger in order to deliver shocks to the ferret's tongue. All procedures for behavioral testing of ferrets were approved by the institutional animal care and use committee (IACUC) of the University of Maryland, College Park.

C.   Results: Ferrets can automatically fuse harmonic partials

Results from testing 3 female ferrets, in Figure 3, demonstrate that ferrets can automatically distinguish harmonic complex tones from inharmonic complex tones. The figure shows performance on consecutive sessions after a training period lasting 15 to 75 sessions had been completed. Hit rates greater than 70% and false-alarm rates less than 20% show that all three ferrets attained proficiency at the baseline task of distinguishing pure-tone targets from inharmonic-complex-tone references. Two ferrets (top two panels) out of three also detected harmonic-complex-tone probes at a significantly higher rate than the false-alarm rate, especially in the first probe sessions. Therefore the ferrets automatically, without training, heard harmonic complex tones as being different than

7

inharmonic complex tones; this finding is the main result of the experiment.

[Insert Figure 3 Here]

D. Discussion

It is worth noting three points from the results. First, two ferrets heard harmonic complex tones to be different from inharmonic complex tones. More generally, they confused harmonic probes with pure-tone targets by responding to them as if they were pure tones[2]. Because there are no consistent differences in timbre between the harmonic and inharmonic tones, the most likely perceptual dimension along which the ferrets categorized these stimuli is fusion. Therefore, we conclude that ferrets, like humans, automatically fuse partials of harmonic complex tones.

Second, the failure of ferret 3 to detect probes at a higher rate than the false-alarm rate does not mean that it could not hear a difference between harmonic and inharmonic complex tones.

---

[2]A possible objection to this interpretation is that the ferrets did not confuse the harmonic tones with the pure tones but simply heard them as a new kind of sound and in confusion responded by withdrawing from the reward spout. There are two reasons why this objection might not hold. First, the parameters of the harmonic complex tones were matched in almost every way to those of the inharmonic complex tones. Thus, hearing harmonic probes as a new category of sounds makes the point that the ferrets could hear them to be different from inharmonic tones, and the logical perceptual dimension for the distinction is harmonic fusion. Second, the novelty response should have declined rapidly as the ferret became accustomed to the harmonic probes. However, the elevated probe hit rate persisted for several sessions, especially in ferret 2.

We attempted to verify that novelty did not cause the probe response by repeating the experiment in ferret 2 with the roles of inharmonic tones and pure tones reversed; i.e., inharmonic tones were targets and pure tones were references. The results of this experiment were inconclusive. Cues available for distinguishing the target from the references were differences in perceptual fusion and differences in timbre. The ferret successfully learned to make this distinction, detecting the target at a hit rate greater than 70%. When harmonic probes were introduced, failure to detect them would have indicated that the harmonic complex tones were perceptually fused, thus making it unlikely that novelty explained the probe response in the original experiment. However, during the probe sessions, the hit rate for the probes was not significantly different than the hit rate for the targets. As in the original experiment, this result does not mean that the ferret did not perceptually fuse the harmonic tones. It is probable that the ferret learned to rely mainly on the timbre difference between reference and target in the baseline task, and therefore classified the harmonic probes as "warning" or target sounds.

The ferret might have learned to focus on the timbre cue in order to perform the baseline task. During probe sessions, it might have continued to use the same cue and thus correctly categorized the harmonic probes to be similar to the inharmonic references along the perceptual dimension of timbre, while ignoring audible differences in fusion.

Finally, the probe hit rate for ferret 1 declined steadily after the first probe session. A probable reason is that the ferret, realizing gradually that the harmonic probes were not associated with an aversive stimulus, adjusted its judgments to use timbre exclusively rather than timbre in conjunction with fusion. Ferret 2, on the other hand, might not have adjusted its judgments simply because it was slower to learn that harmonic probes were not associated with the aversive shock. Given more probe sessions, the ferret might have exhibited such a learning effect. Indeed, consistent with the notion that learning ability might underlie the difference between the declining probe hit rate for ferret 1 and the lack of such decline for ferret 2, ferret 2 took almost five times more sessions than ferret 1 to learn the baseline task.

III.   Experiment 2: Correlates of harmonic fusion in primary auditory cortical neurons

A.   Rationale

To seek correlates of harmonic fusion, we recorded single-unit activity from primary auditory cortex (AI) to a sequence of harmonic and inharmonic complex toneswhere all the sounds in the sequence shared a component at the best frequency (*BF*) of the unit under investigation (Figure 4). The frequency of this shared partial is known as the *anchor frequency* (*AF*) and the sequence is known as an *anchored tone sequence* in the rest of the paper. The anchor component was placed at the *BF* to ensure a robust response from the neuron for every complex tone. Any special cortical or upstream (subcortical) neurons computation on harmonic sounds would be expected to result in a systematic difference in the responses to the harmonic complex tones in the sequence compared to those for the inharmonic complex tones.

[Insert Figure 4 Here]

The spectrotemporal tuning characteristics of cortical neurons ((Schreiner and Calhoun, 1994; Shamma, Versnel, and Kowalski, 1995; Kowalski, Depireux, and Shamma, 1996; Depireux et al., 2001)) can cause an effect of the spectrotemporal integration of stimulus energy (e.g., due to interaction of inhibitory sidebands with balance of spectrum above and below *BF*) to be confused with an effect of harmonic context. To account for spectrotemporal filtering, we measured a neuron's spectrotemporal receptive field (*STRF*) and used it to predict the responses to the stimuli in the anchored tone sequences. An effect of harmonic context should then appear as a systematic difference in the predictability of the responses to harmonic complex tones compared to those for inharmonic complex tones, because the predictions using the *STRF* account for the effect of the spectral context. The *STRF* is the best-fitting linear model for the transformation of the stimulus spectrotemporal envelope by a neuron. Because we measured *STRF*s using spectrotemporal envelopes imposed upon broadband noise (inharmonic) carriers, an effect of harmonic fusion is expected to manifest itself as a reduction in the predictability of responses to harmonic tones compared to that for inharmonic tones. A reduction is expected because the *STRF*, by definition, gives the best possible linear estimate of the response; any modification due to harmonicity must result in a degradation of this best estimate.

B.   Methods

1.   Experimental apparatus and methods

Animal preparation and recording procedures, which were approved by the IACUC of the University of Maryland, are described in detail in another publication (Fritz et al., 2003). We give a brief outline here. Ferrets were adapted to lie motionless in a custom apparatus that restrained them. The auditory cortex was accessed through a small hole of diameter less than 0.5-mm with a parilyne-coated tungsten microelectrode having resistance ranging from 2 to $7M\Omega$ at 1 kHz. Only one hole in the skull was used at a time in recording sessions of 4 to 6 hours, with precautions taken to maintain sterility of the hole at all times. Each hole was used for $5 - 7$ days after which

it was sealed with dental cement. The small size of the hole afforded stable recordings and lowered chances of infection. Closely spaced holes were made over auditory cortex, with particular attention to the low-frequency areas for this study. Recordings were attributed to AI based on tonotopic organization and response properties (strong response to tones and relatively short response latency), but a few penetrations might have been from adjacent areas.

Electrode penetrations were made perpendicular to the surface of the cortex with a hydraulic microdrive under visual guidance via an operating microscope. Recordings yielded spikes from $1-3$ neurons that were sorted offline with a combination of an automatic spike-sorting algorithm (Lewicki, 1994) and manual techniques. A spike class was included as a single unit for further analysis if less than 95% of interspike intervals were smaller than 1 ms, the putative absolute refractory period.

Sounds were delivered with an Etymotic ER-2 earphone inserted into the entrance of the ear canal on the contralateral side of the cortex being investigated. All stimuli were generated by computer and fed through equalizers to the earphone. An Etymotic ER-7C probe-tube microphone was used to calibrate the sound system *in situ*. An automatic calibration procedure gave flat frequency responses below 20000 Hz.

2. **Stimuli and analysis**

Stimuli   After a cluster of single units was isolated using pure-tone and complex-tone search stimuli, its response area was measured with pure tones to get its *BF*. The change of discharge rate as a function of stimulus level was measured with *BF* pure tones at a range of levels in order to estimate the threshold. Responses to a tone sequence anchored at the *BF* (as in Figure 4) were measured at approximately 20 dB above *BF*-tone threshold. Finally, the spectrotemporal filtering characteristics were characterized at the same level as the anchored-tone sequence by measuring the *STRF* with temporally-orthogonal-ripple-combination (*TORC*) stimuli (Klein et al., 2000).

Anchored tone sequences   Anchored tone sequences consisted of 6-component harmonic and inharmonic complex tones. Up to six different harmonic complex tones with components in random

11

phase were part of the sequence, where each tone had a different component number at the anchor frequency. For most units, the second through fifth components were at the anchor frequency but a few had the first through fifth or all six component numbers at the anchor frequency. Because no differences were observed between these cases, we do not distinguish them in the presentation of results.

Inharmonic complex tones in the sequences were formed by scrambling ratios between consecutive partials of a harmonic complex tone, as in Experiment 1. The same sequence of ratios were used in each of the four or five different inharmonic complex tones for almost all units. In some of the early recording sessions (4 of the reported single units), a few inharmonic complex tones in the sequence were formed in three additional ways:

1. All partials of a harmonic tone were additively shifted by a fixed amount.

2. The anchor component was shifted by 10% in an otherwise harmonic complex tone.

3. The anchor component was shifted by 10% in a complex tone formed by additively shifting the partials of a harmonic tone by a fixed amount.

These cases did not yield different results than the later recording sessions, so we do not distinguish them in the presentation of results.

All components of the harmonic and inharmonic complex were presented at the same level. The pure tone in the sequence also had the same level as individual components of the complex tones, so that its overall intensity was less than that of the complex tones.

Characterizing linear processing of spectrotemporal envelopes with *STRF*s   Underlying the measurement of a *STRF* is the observation ((Schreiner and Calhoun, 1994; Shamma, Versnel, and Kowalski, 1995; Kowalski, Depireux, and Shamma, 1996)) that responses of AI neurons have a large linear component with respect to the spectrotemporal envelope of sounds. The $STRF(t, x)$, a two-dimensional function of time t and log frequency $x = log(f/f_0)$, describes the linear component of the transformation between the spectrotemporal envelope of an acoustic stimulus and the

neural response. This response component is given by

$$r(t) = \int \int STRF(\tau, x) \cdot S(t - \tau, x) \, d\tau \, dx \qquad (1)$$

where $r(t)$ is the instantaneous discharge rate of the neuron and $S(t, x)$ is the spectrotemporal envelope (or dynamic spectrum) of the stimulus; the equation describes a convolution in time and a correlation in log frequency. Intuitively, the response of the neuron r at time t is the correlation of the *STRF* with the time-reversed dynamic spectrum of the stimulus S around that moment. This operation can be viewed as similar to a matched-filtering operation whereby the maximum response of the neuron occurs when the time-reversed dynamic spectrum most resembles the *STRF*.

The theory and practice of measuring *STRF*s with *TORC* stimuli are in Klein et al (2000). A brief outline is given in the appendix. *TORC*s are composed of *moving ripples* (Kowalski, Depireux, and Shamma, 1996; Depireux et al., 2001), which are broadband sounds having sinusoidal temporal and spectral envelopes. Moving ripples are basis functions of the spectrotemporal domain in that arbitrary spectrotemporal envelopes can be expressed as combinations of these stimuli. The accumulated phase-locked responses to individual moving ripples gives a spectrotemporal modulation transfer function (parameterized by ripple velocity or temporal modulation rate and ripple density or spectral modulation rate) whose two-dimensional inverse Fourier transform is the *STRF*. *TORC*s are special combinations of moving ripples such that two components having different ripple densities do not share the same ripple velocity. This special combination of moving ripples enables rapid measurement of the *STRF*. We used tones densely spaced on a log-frequency axis (100 tones/octave, spanning 5 octaves) and in random phase as carriers for the envelope of the *TORC* stimuli.

It is worth noting two properties of our *STRF* measurements.

1. Underlying our measurement of the *STRF* is an assumption that the neuronal system has reached a sinusoidal steady-state. Consequently, the *STRF* quantifies changes of the discharge rate above and below a steady-state rate.

2. The *STRF*s are zero-mean because the responses to static ripples (moving ripples with ripple velocity of 0 Hz) are defined to be zero (Depireux et al., 2001). Therefore, the *STRF* does

not predict any response to the steady-state part of static sounds[3].

Quantifying the effects of neural filtering on responses to complex tones   In order to account for the effect of spectrotemporal filtering on responses of a unit to the anchored tone sequence, the measured *STRF* was used to predict the peri-stimulus time (PST) histogram of the response to each stimulus. The predictability of the PST histogram then served as a measure of the response to each stimulus with spectrotemporal filtering removed. Correlates of harmonic fusion were expected to show up as systematic differences in the predictabilities of the response for harmonic stimuli compared to those for inharmonic stimuli.

A discrete-time implementation of Equation 1 was used to predict the response to each stimulus in the anchored tone sequence (visualized in Figure 5). The predicted PST histogram ($p[n]$) for a stimulus was a function of the sum of two terms, one term due to spectrotemporal filtering ($p_{strf}[n]$) and a second term representing the steady-state discharge rate (p_{ss}). For a complex tone having L components, $p_{strf}[n]$ was obtained in two steps.

1. Convolve trapezoid-like stimulus envelope (10 ms cosine-squared onset and offset ramps) $e[n]$ with horizontal slices of the *STRF* at the frequencies x^i of the partials, $STRF[n, x^i]$, to get the contribution of individual partials to the overall prediction.

$$p_{strf}^i[n] = \frac{1}{N}(STRF[n, x^i] * e[n]) \tag{2}$$

   N is the length of $STRF^i[n]$ and $*$ indicates convolution in time.

2. Combine contributions of individual partials to get the overall prediction from the *STRF* as the mean of the individual contributions.

$$p_{strf}[n] = \frac{1}{L} \sum_{i=1}^{L} p_{strf}^i[n] \tag{3}$$

---

[3]In order to predict the steady-state discharge rate, it is necessary to measure the DC-gain terms of the *STRF*; i.e., the responses to static ripples. These cannot be measured by directly incorporating static ripples into *TORC* stimuli because responses to the static ripples are difficult to disambiguate from nonlinear responses to the moving ripples resulting from, for example, saturation and rectification. The DC-gain terms can be estimated separately using static ripples presented at different stimulus levels.

14

Finally, the predicted PST histogram was

$$p[n] = g(p_{strf}[n] + p_{ss}) \tag{4}$$

where $g(\cdot)$ indicates half-wave rectification. Because the *STRF* does not predict the steady-state discharge rate, we used the measured steady-state discharge rate r_{ss} instead of p_{ss}. In Figure 5D, shading is used to indicate one standard deviation above and below the mean prediction, where the standard deviation is estimated by resampling the measured *STRF* using the bootstrap technique (Efron and Tibshirani, 1993). Figure 5C and E show raster and PST histograms of the actual response of the cell to the same stimulus, demonstrating the relatively high quality of the prediction in this case. Shading in the PST histogram (Figure 5E indicates one standard deviation above and below the mean, where the statistics were obtained by bootstrap resampling the response.

[Insert Figure 5 Here]

The steady-state discharge rate is treated separately because our measurement of the *STRF* does not directly predict it. However, it is clear from the frequency tuning of cortical neurons that their steady-state responses are influenced by the stimulus spectrum. Because the predictability of the steady-state discharge rate (as quantified in this paper)[4] did not lead to different conclusions than

---

[4]We account for the effect of frequency tuning on the stimulus spectrum by

1. Extracting the spectral slice $RF[x]$ associated with the largest singular value from a singular value decomposition of the *STRF*

2. Using $RF[x]$ to get the relative magnitudes of the responses to each stimulus in the anchored tone sequence as

$$\psi_j = \sum_{i=1}^{L} RF[x^i]$$

   where x^i are the component frequencies of the stimulus, L is the number of components, and the subscript for ψ indicates the j-th stimulus in the anchored tone sequence.

The absolute magnitude of the responses cannot be predicted with this procedure. In order to compare actual discharge rates with predictions of the same order of magnitude, we scaled predictions such that the largest ψ_j had the same magnitude as the largest steady-state discharge rate; i.e., the scaled prediction of the steady-state discharge rate is

$$\Psi_j = \psi_j \cdot \frac{\max_j[rss_j]}{\max_j[\psi_j]}$$

15

those due to the predictability of other response components, we do not report the results of the analysis in any detail.

C.   Results: A small fraction of primary auditory cortical neurons distinguish harmonic complex tones from inharmonic complex tones

We report data from 34 single units in 4 ferrets. The *BF*s of these units varied from 210 Hz to 8000 Hz, although only four had *BF* greater than 1500 Hz.

Figure 6 shows the responses of a single unit to the harmonic and inharmonic tones constituting the anchored tone sequences, where harmonic tones are labeled with the prefix 'H' and the inharmonic tones are labeled with the prefix 'SR'. The spectral context clearly influenced the responses of this unit to the different tones. For example, the harmonic tones labeled H3 and H4 (3rd and 4th partials at the *BF* respectively) were inhibited for the first half of stimulus presentation, while none of the other stimuli elicited such a response. Similarly, the unit responded far more weakly to the inharmonic tones labeled SR3 and SR4 than to any of the other stimuli[5].

[Insert Figure 6 Here]

In order to account for the effect of spectrotemporal filtering in this unit, we used the *STRF* to predict PST histograms for the complex tones. Figure 7A shows actual PST histograms overlayed on predictions for some of the stimuli. Predictions were normalized by the maximum across

Such prediction is suited for comparing the predicted pattern of variation of the rate across stimuli with the pattern actually obtained, but it is not valid for comparing the absolute magnitudes of the predictions with those of the actual responses.

[5]Interestingly for this unit, the response to stimulus SR1 shows temporal structure in the discharge pattern. This temporal structure reflects synchronization of discharges to envelope modulations resulting from interaction between tone components. Responses to all stimuli showed such synchronization to tone interactions up to 250 Hz. As a result of this 250 Hz upper limit, the temporal envelope for harmonic tones H1 through H4 was not reflected in the neural discharge patterns but the temporal envelope for H5 did produce synchronized discharges. The rate limit also makes such units unsuitable for encoding the periodicity pitch of harmonic complex tones; instead, they are better suited for encoding residue pitch. A similar limit for synchronizing to the stimulus temporal envelope has been observed in previous studies of auditory cortex (Steinschneider et al., 1994; Liang, Lu, and Wang, 2001; Schulze et al., 2002; Elhilali et al., 2004). No other unit in our population exhibited such temporal discharge patterns.

all prediction waveforms and actual PST histograms were normalized by the maximum across all actual responses, because the magnitudes of the predictions were consistently greater than the magnitudes of the PST histograms and this normalization better reveals temporal patterns of discharge. The response was better predicted for the inharmonic stimuli than for the harmonic stimuli, mainly because the *STRF* was unable to predict the late response for stimuli H3 and H4. We summarized the linear predictability of the response with a correlation coefficient (ρ) between the response PST histogram and the predicted waveform[6]. Figure 7C shows distributions of ρ for the harmonic tones pooled together and the inharmonic tones pooled together, with bootstrap resamples included in the distributions. The distributions show that ρ tends to be lower for harmonic stimuli than for inharmonic stimuli, reflecting the lower predictability of responses to the harmonic stimuli. Nevertheless, there is substantial overlap in the distributions of ρ for harmonic and inharmonic tones because differences in the responses to these classes of stimuli were dominated by a few of the harmonic stimuli. Therefore, this unit distinguished the harmonic tones from the inharmonic tones to some extent.

[Insert Figure 7 Here]

Most units in the population either distinguished the harmonic tones from the inharmonic tones even more weakly, or failed to do so at all. Figure 8A shows predictions and PST histograms for some stimuli from such a unit. This neuron tended to respond at the onset of stimuli. These responses were predicted well by the *STRF*, and the histograms of correlation coefficients in Figure 8C show that the predictability of the response did not differ greatly for harmonic tones and inharmonic tones.

[Insert Figure 8 Here]

In order to summarize, for any given unit, whether the predictability of the responses for har-

---

[6]The correlation coefficient between the response $r[n]$ and the prediction $p[n]$ is

$$\rho = \frac{\sum_n r[n] \cdot p[n] - \overline{r} \cdot \overline{p}}{\sigma_r \cdot \sigma_p}$$

where $\overline{r}$ and $\overline{p}$ are the mean response and mean prediction, while σ_r and σ_p are the standard deviations of the response and the prediction respectively. Because the covariance in the numerator is normalized by the standard deviations, ρ is insensitive to the overall magnitudes of $r[n]$ and $p[n]$.

17

monic tones were collectively different than those for inharmonic stimuli, we quantified the difference between the histogram for harmonic stimuli and that for inharmonic stimuli with a distance d_ρ. If $hist_h[n]$ and $hist_{ih}[n]$ are histograms of ρ for harmonic and inharmonic stimuli, then

$$d_\rho = \sum_n \left| \frac{hist_h[n]}{\sum hist_h[n]} - \frac{hist_{ih}[n]}{\sum hist_{ih}[n]} \right| \tag{5}$$

where the two fractions within the summation convert the histograms into probability mass functions and $|\cdot|$ is the absolute value operation. Two identical histograms have the minimum distance of 0 and two histograms that do not overlap at all have the maximum distance of 2. Figure 9 shows the distributions of the distances in the population of 34 units, with the square on the abscissa indicating d_ρ for the unit in Figure 6 and 7 and the circle indicating d_ρ for the unit in Figure 8. Most units had distances lower than that of Figure 6 and 7. Distances, typically, were similar to that of Figure 8. Units having d_ρ of 1 or greater could be said to distinguish harmonic tones from inharmonic tones, and these constituted only 5 out of 34 units in the population. None of these 5 units distinguished harmonic tones from inharmonic tones robustly.

[Insert Figure 9 Here]

The correlation coefficient quantifies the positions and relative amplitudes of peaks and valleys in the predicted waveforms and the PST histograms, but it does not quantify how well the magnitude of the predicted waveform compares to the magnitude of the time-varying component of PST histograms. Furthermore, the steady-state discharge rate is not predicted by the *STRF* at all (it is defined to be 0), but it must be affected by the interaction of the stimulus spectrum with the frequency receptive field. Accounting for these factors that are not quantified by the correlation coefficient (not shown) did not alter the conclusions based on the correlation coefficient. Roughly 10% of the units expressed correlates of harmonicity when the predictability of these other response components was examined, and these correlates were weak.

D.   Discussion

About 10% of units in AI showed some differences between their responses to harmonic and inharmonic complex tones. Given how readily harmonic sounds are perceived as unitary entities by

humans (von Helmholtz, 1863) and by ferrets (Experiment 1), these units are unlikely to underlie harmonic fusion. These results are consistent with previous studies indicating that periodicity pitch is not computed in AI.

One caveat to the conclusion is that a significant fraction of units in the population, approximately 2/3, were transient responders in that either the onset or offset responses were vigorous but steady-state discharge rate was not significantly different than the spontaneous discharge rate. As a result, if the expression of harmonic fusion is subtle and occurs a few hundred milliseconds after stimulus onset, the absence of significant steady-state activity may have masked differences in the responses to harmonic and inharmonic stimuli. Imposing slow amplitude modulation below 30 Hz upon the complex tones might help reveal such differences by driving neurons more vigorously (Schreiner and Urbas, 1988; Kowalski, Depireux, and Shamma, 1996; Liang, Lu, and Wang, 2001; Depireux et al., 2001; Eggermont, 2002; Elhilali et al., 2004) without significantly altering perceptual fusion (Darwin, Ciocca, and Sandell, 1994).

The harmonic complex tones in the anchored tone sequence resemble the stimuli used by Fishman et al (2000), who measured multi-unit activity and current-source density profiles in AI of monkeys in response to complex tones consisting of 3 consecutive harmonics, with the middle component always at the BF of the recording location. Discharge rates were predictable from the pure-tone tuning curve when the middle component was approximately greater than component number 5 but not when it was less. This finding was attributed to an effect of resolvability of individual harmonics upon the predictability of the response. We failed to find such an effect of harmonic number. One possible reason for the discrepancy is the use of low harmonic numbers in our stimuli, always less than component number 6. A second possible reason is that the STRF (a dynamic measure of neural tuning) is used for prediction in the present study as opposed to the pure-tone tuning curve (a static measure of neural tuning) used by Fishman et al (2000).

IV.   General discussion

Our results demonstrate that ferrets hear harmonic complex tones as fused, unitary entities but that this fusion does not leave its imprint upon neurons in AI. The absence of cortical neural correlates of fusion is consistent with conclusions of previous studies that direct correlates of periodicity pitch do not exist in AI.

Using harmonic complex tones with and without the fundamental component, Schwarz and Tomlinson (1990) failed to find any neurons in monkey AI that responded as if they computed the periodicity pitch. Similarly, Steinschneider et al (1998) used alternating and uniform polarity click trains that allow residue pitch due to waveform periodicity to be distinguished from periodicity pitch due to stimulus spectrum, and failed to find any multi-unit activity that correlated directly with periodicity pitch. Recent neuromagnetic and magnetic resonance imaging studies also support the conclusion that periodicity pitch is not extracted by AI. Instead, investigations with regular-interval noise and filtered harmonic complex tones suggest that pitch due to spectrally resolved harmonics is computed in an area that is anterolateral to Heschl's gyrus, the locus of AI in humans (Krumbholz et al., 2003; Patterson et al., 2003; Penagos, Oxenham, and Melcher, 2003).

Our results pose a challenge to models that posit the central nucleus of the inferior colliculus (IC) or lower brainstem nuclei as loci of an across-frequency integration step required for computing periodicity pitch (e.g. Langer, 1992; Shamma and Klein, 2000). AI is part of the core or lemniscal auditory pathway originating in central IC and characterized by short-latency sharply-tuned tone responsive units, as opposed to the belt pathway which originates in non-central areas of IC (Andersen et al., 1980; Andersen, Snyder, and Merzenich, 1980; Winer, 1992). Extraction of periodicity pitch anywhere along the lemniscal pathway should have been evident in our data.

The problem of identifying the cortical neural correlates for the special perceptual status of harmonic sounds remains a vexing one. Recent functional imaging studies suggesting that periodicity pitch is extracted in non-primary cortical fields (Krumbholz et al., 2003; Patterson et al., 2003; Penagos, Oxenham, and Melcher, 2003; Warren et al., 2003) provide guidance on the cortical areas that should be investigated in future single-unit studies.

V. Acknowledgments

A Measuring the *STRF* of a neuron with *TORC* stimuli

The *TORC* stimulus is a particular combination of broadband stimuli known as *moving ripples*, whose spectrotemporal envelope is given by

$$S(t, x) = a_0 + a \cos[2\pi(\omega t + \Omega x) + \psi] \tag{6}$$

At each frequency location, the function describes a sinusoidal modulation of intensity at a rate of ω cycles/second around a mean a_0 and amplitude a; the relative phases of these modulations at different x produce a sinusoidal or rippled profile of density Ω cycles/octave. The rippled profile drifts across the spectral axis in time, hence leading to the name of moving ripples for these stimuli. Analogous to sinusoids for one-dimensional signals, moving ripples are basis functions of the spectrotemporal domain in that any arbitrary spectrotemporal profile can be constructed from a combination of them. And similarly analogous to estimating the impulse response of a one-dimensional system using reverse correlation with white noise stimuli that have equal representation of all sinusoidal frequencies within the system bandwidth (de Boer and de Jongh, 1978), it is possible to estimate the *STRF* by reverse correlation with spectrotemporal white noise ($STWN$), which is a stimulus that has an equal representation of all moving ripples within the spectrotemporal bandwidth of the system[7] (Klein et al., 2000; Klein et al., 2003).

However, because the $STRF(x, t)$ transforms a 2-dimensional input to a 1-dimensional output, moving-ripple components of $STWN$ that are spectrally orthogonal (different ripple densities, Ω, but same ripple velocities, ω) can evoke overlapping response components that cannot be dis-

---

[7]For the cortical neurons investigated in this paper, ripple velocities were less than 32 Hz and ripple densities were less than 1.4 cycles/octave.

ambiguated; reverse correlation with such stimuli can result in inaccurate and noisy estimates of the *STRF*. The *TORC* stimulus overcomes this problem by ensuring that each moving ripple in the stimulus has a different absolute ripple velocity $|\omega|$, so that each linear response component is uncorrelated with every ripple component of the stimulus except for the one evoking it. Therefore, when using reverse correlation with a *TORC* stimulus, response components of a given ripple velocity contribute only to a single $[\omega, \Omega]$ component of the *STRF*.

References

Andersen, R., Roth, G., Aitkin, L., and Merzenich, M. (1980). "The efferent projections of the central nucleus and the pericentral nucleus of the inferior colliculus in the cat," J. Comp. Neurol. **194**, 649–662.

Andersen, R., Snyder, R., and Merzenich, M. (1980). "The topographic organization of cortico-collicular projections from physiologically identified loci in the AI, AII, and anterior auditory cortical fields of the cat," J. Comp. Neurol. **191**, 479–494.

Bregman, A. S. (1990). *Auditory Scene Analysis* (MIT Press, Cambridge, MA).

Brokx, J. and Nooteboom, S. (1980). "Intonation and perceptual separation of simultaneous voices," J. Phonetics **10**, 23–36.

Brunstrom, J. M. and Roberts, B. (1998). "Profiling the perceptual suppression of partials in periodic complex tones: Further evidence for a harmonic template.," J. Acoust. Soc. Am. **104**, 3511–9.

Brunstrom, J. M. and Roberts, B. (2000). "Separate mechanisms govern the selection of spectral components for perceptual fusion and for the computation of global pitch.," J. Acoust. Soc. Am. **107**, 1566–77.

Ciocca, V. and Darwin, C. (1993). "Effects of onset asynchrony on pitch perception: adaptation or grouping?," J. Acoust. Soc. Am. **93**, 2870–8.

Darwin, C. and Ciocca, V. (1992). "Grouping in pitch perception: Effects of onset asynchrony and ear of presentation of a mistuned component," J. Acoust. Soc. Am. **91**, 3381–3390.

Darwin, C., Ciocca, V., and Sandell, G. J. (1994). "Effects of frequency and amplitude modulation on the pitch of a complex tone with a mistuned harmonic," J. Acoust. Soc. Am. **95**, 2631–6.

de Boer, E. (1956), "On the "residue" in hearing," Ph.D., University of Amsterdam.

de Boer, E. (1976). "On the "residue" and auditory pitch perception," in *Handbook of Sensory Physiology*, edited by W. Keidel and W. Neff (Springer Verlag, Berlin), Vol. 3, pp. 479–583.

de Boer, E. and de Jongh, H. R. (1978). "On cochlear encoding: Potentialities and limitations of the reverse-correlation technique," J. Acoust. Soc. Am. **63**, 115–135.

Depireux, D. A., Simon, J. Z., Klein, D. J., and Shamma, S. A. (2001). "Spectro-temporal response field characterization with moving ripples in ferret primary auditory cortex," J. Neurophys. **85**, 1220–34.

Efron, B. and Tibshirani, R. (1993). *An Introduction to the Bootstrap* (Chapman and Hall, New York).

Eggermont, J. (2002). "Temporal modulation transfer functions in cat primary auditory cortex: Separating stimulus effects from neural mechanisms," J. Neurophys. **88**, 305–321.

Elhilali, M., Fritz, J., Klein, D., Simon, J., and Shamma, S. (2004). "Dynamics of precise spiking in primary auditory cortex," J. Neurosci. **24**, 1159–72.

Fishman, Y., Reser, D. H., Arezzo, J. C., and Steinschneider, M. (2000). "Complex tone processing in primary auditory cortex of the awake monkey. II. Pitch versus critical band representation," J. Acoust. Soc. Am. **108**, 247–62.

Flanagan, J. L. and Gutman, N. (1960). "Pitch of periodic pulses without fundamental component," J. Acoust. Soc. Am. **32**, 1319–1328.

Fritz, J., Shamma, S., ElHilali, M., and Klein, D. (2003). "Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex," Nat. Neurosci. **6**, 1216–23.

Goldstein, J. (1973). "An optimum processor theory for the central formation of the pitch of complex tones," J. Acoust. Soc. Am. **54**, 1496–1516.

Hartmann, W., McAdams, S., and Smith, B. (1993). "Hearing a mistuned harmonic in an otherwise periodic complex tone," J. Acoust. Soc. Am. **88**, 1712–24.

Heffner, H. and Whitfield, I. (1976). "Perception of the missing fundamental by cats," J. Acoust. Soc. Am. **59 No.4**, 915–919.

Heffner, H. E. and Heffner, R. S. (1995). "Conditioned avoidance for animal psychoacoustics," in *Comparative Psychoacoustics*, edited by xx (xx, xx), pg. xx.

Houtsma, A. and Smurzynski, J. (1990). "Pitch identification and discrimination for complex tones with many harmonics," J. Acoust.Soc.Am. **87**, 304–310.

Kadia, S. and Wang, X. (2003). "Spectral integration in A1 of awake primates: Neurons with single- and multipeaked tuning characteristics," J. Neurophys. **89**, 1603–22.

Klein, D. J., Depireux, D. A., Simon, J. Z., and Shamma, S. A. (2000). "Robust spectro-temporal reverse correlation for the auditory system: Optimizing stimulus design," J. Comput. Neurosci. **9**, 85–111.

Klein, D. J., Depireux, D. A., Simon, J. Z., and Shamma, S. A. (2003). "Linear stimulus-invariant processing and spectrotemporal reverse correlation in primary auditory cortex," J. Comput. Neurosci. **Submitted**.

Kowalski, N., Depireux, D., and Shamma, S. A. (1996). "Analysis of dynamic spectra in ferret primary auditory cortex. I. Characteristics of single unit responses to moving ripple spectra," J. Neurophys. **76**, 3503–3523.

Krumbholz, K., Patterson, R., Seither-Preisler, A., Lammertmann, C., and Lutkenhoner, B. (2003). "Neuromagnetic evidence for a pitch processing center in heschl's gyrus," Cerebral Cortex **13**, 765–772.

Langner, G. (1992). "Periodicity coding in the auditory system," Hearing Res. **60**, 115–142.

Langner, G., Sams, M., Heil, P., and Schulze, H. (1997). "Frequency and periodicity are represented in orthogonal maps in the human auditory cortex: Evidence from magnetoencephalography," J. Comp. Physiol. [A] **181**, 665–76.

Lewicki, M. (1994). "Bayesian modeling and classification of neural signals," Neural Computation **6**, 1005–30.

Liang, L., Lu, T., and Wang, X. (2001). "Neural representations of sinusoidal amplitude and frequency modulations in the primary auditory cortex of awake primates," J. Neurophys. **87**, 2237–61.

Lin, J. and Hartmann, W. (1998). "The pitch of a mistuned harmonic: Evidence for a template model," J. Acoust. Soc. Am. **103**, 2608–17.

McAlpine, D. (2002). "Are pitch neurones the result of difference tones on the basilar membrane?," Assoc. Res. Otolaryngol. Abstr. **25**, 154.

Moore, B. and Glasberg, B. R. (1985). "Relative dominance of individual partials in determining the pitch of complex tones," J. Acoust. Soc. Am. **77**, 1853–1860.

Moore, B., Peters, R., and Glasberg, B. (1986). "Thresholds for hearing mistuned partials as separate tones in harmonic complexes," J. Acoust. Soc. Am. **80**, 479–483.

Moore, D. and Kitzes, L. (1985). "Projections from the cochlear nucleus to the inferior colluculus in normal and neonatally cochlea-ablated gerbils.," J. Comp. Neurol. **240**, 180–195.

Patterson, R., Uppenkamp, S., Johnsrude, I., and Griffiths, T. (2003). "The processing of temporal pitch and melody information in auditory cortex," Neuron **36**, 767–776.

Penagos, H., Oxenham, A., and Melcher, J. (2003). "Effects of harmonic resolvability on the cortical activity produced by complex tones," Assoc. Res. Otolaryngol. Abstr. **26**, 1125.

Schreiner, C. and Urbas, J. (1988). "Representation of amplitude modulation in the auditory cortex of the cat: Ii comparison between cortical fields," Hearing Res. **32**, 49–64.

Schreiner, C. E. and Calhoun, B. M. (1994). "Spectral envelope coding in the cat primary auditory cortex: Properties of ripple transfer functions," Aud. Neurosci. **1**, 39–61.

Schulze, H., Hess, A., Ohl, F., and Scheich, H. (2002). "Superposition of horseshoe-like periodicity and linear tonotopic maps in auditory cortex of the Mongolian gerbil," Eur. J. Neurosci. **15**, 1077–1084.

Schulze, H. and Langner, G. (1997a). "Periodicity coding in the primary auditory cortex of the Mongolian gerbil (Meriones unguiculatus): Two different strategies for coding pitch and rhythm," J. Comp. Physiol. [A] **181**, 651–63.

Schulze, H. and Langner, G. (1997b). "Representation of periodicity pitch in the primary auditory cortex of the Mongolian gerbil," Acta Otolaryngol Suppl. **532**, 89–95.

Schwarz, D. W. and Tomlinson, R. W. (1990). "Spectral response patterns of auditory cortex neurons to harmonic complex tones in alert monkey (macaca mulatta).," J Neurophysiol **64**, 282–98.

Shamma, S. and Klein, D. J. (2000). "The case of the missing pitch templates: How harmonic templates emerge in the early auditory system," J. Acoust. Soc. Am. **107**, 2631–44.

Shamma, S. A., Versnel, H., and Kowalski, N. (1995). "Ripple analysis in ferret primary auditory cortex. I. Response characteristics of single units to sinusoidally rippled spectra," Aud. Neurosci. **1**, 233–254.

Steinschneider, M., Reser, D., Fishman, Y., Schroeder, C., and Arezzo, J. (1994). "Click train encoding in primary auditory cortex of the awake monkey: Evidence for two mechanisms subserving pitch perception," J. Acoust. Soc. Am. **104**, 2935–2955.

Terhardt, E. (1974). "Pitch, consonance, and harmony," J. Acoust. Soc. Am. **55**, 1061–1069. virtual pitch, place.

Terhardt, E. (1979). "Calculating virtual pitch," Hearing Res. **1**, 155–182.

Tomlinson, R. W. W. and Schwartz, D. W. F. (1988). "Perception of the missing fundamental in nonhuman primates," J. Acoust. Soc. Am. **84**, 560–565.

von Helmholtz, H. (1863). *Die Lehre von den Tonempfindungen als physiologische Grundlage fr die Theorie der Musik.* (Vieweg und Sohn, Braunschweig).

Warren, J. D., Uppenkamp, S., Patterson, R. D., and Griffiths, T. D. (2003). "Separating pitch chroma and pitch height in the human brain," Proc. Nat. Acad. Sci. **100**, 10038–42.

Whitfield, I. C. (1980). "Auditory cortex and the pitch of complex tones," J. Acoust. Soc. Am. **67**, 644–647.

Wightman, F. L. (1973). "The pattern-transformation model of pitch," J. Acoust. Soc. Am. **54**, 407–416.

Winer, J. A. (1992). "The functional architecture of the medial geniculate body and the primary auditory cortex," in *The mammalian auditory pathway: Neuroanatomy*, edited by D. B. Webster, A. N. Popper, and R. R. Fay (Springer-Verlag, New York).

Zatorre, R. (1988). "Pitch perception of complex tones and human temporal-lobe function," J. Acoust. Soc. Am. **84**, 566–572.

Tables

Trial type	Spout contact (touch period)	Spout contact (retract/shock period)	Response class
Reference	Contact	No contact	Safe
Reference	Contact	Contact	False alarm
Reference	No contact	N/A	Early
Target	Contact	No contact	Hit
Target	Contact	Contact	Miss
Target	No contact	N/A	Early

TABLE I. Relationship between stimulus timing, ferret response, and performance measures.

Figure Captions

FIG. 1. Stimulus protocols for testing if ferrets automatically fuse partials of harmonic complex tones. (*Top*) In baseline sessions, ferrets are trained to detect pure-tone targets terminating a one- to six-stimulus sequence of inharmonic-complex-tone reference sounds. Cues available for distinguishing the targets from the references are i) differences in the quality of perceptual fusion and ii) differences in timbre. (*Bottom*) In probe sessions conducted after ferrets attain proficiency in baseline sessions, a small fraction of the reference sounds are replaced by harmonic-complex-tone probe sounds. If harmonic complex tones are perceptually fused, then they might occasionally be confused with the pure-tone targets, thereby indicating that ferrets can automatically fuse harmonic partials.

FIG. 2. **A** Schematic representation of a trial sequence, where the target sound is presented on trial 4. The presentation of trials is paused after a target trial until the ferret returns to the spout. **B** Schematic representation of a reference trial shows two time intervals, one before the stimulus onset and the second after stimulus offset, during which the ferret's contact with a reward spout determines the response class for the trial. If the ferret is not in contact with the spout during the first interval, the response is classified as an "early" withdrawal and not counted toward computing overall performance on the experiment. If the ferret fails to contact the spout during the second interval, the trial performance is scored as a "false alarm". **C** Schematic representation of a target trial shows the same two intervals of time. The first interval plays the same role as on a reference trial. The second interval is the period when contact with the spout elicits a shock, and the trial performance is scored as a "miss". Based on a figure from Heffner and Heffner (1995) .

FIG. 3. Performance of three ferrets on baseline and probe sessions of experiment. Several training sessions occurred prior to the first session indicated on the abscissa of the figure for each of the ferrets. All three ferrets attained proficiency at the baseline task of distinguishing pure-tone targets from inharmonic-complex-tone references, as indicated by the hit rate exceeding 70% and the false-alarm rate not exceeding 20%. Two ferrets (top two panels) automatically heard harmonic complex tones as being different from inharmonic complex tones, as indicated by probe hit rates that were significantly greater than the false-alarm rate during the probe sessions.

FIG. 4. Stimulus protocol for investigating auditory cortical correlates of harmonicity in ferrets. Neural activity is recorded for a sequence of complex tones and pure tones (*anchored tone sequence*), all of which share a component at the best frequency (BF) of the neuron under investigation. If cortical (or upstream) neurons treat harmonic sounds in a special manner, then the partial at BF should elicit a qualitatively different response when presented in a harmonic context compared to an inharmonic context.

FIG. 5. Predicting responses to complex tones using *STRF*s. The predicted response (**D**) is the convolution of the spectrotemporal envelope of the stimulus (**A**), schematic spectrogram for complex-tone having 6 partials) with the *STRF* (**B**). More specifically, the operation can be seen as comprising two steps: i) One-dimensional convolutions between the stimulus envelope (trapezoid-like with 10 ms cosine-squared onset and offset ramps) and horizontal slices of the *STRF* corresponding to the frequencies of each partial; ii) Average of the results of the convolutions. These steps for part of an implementation of Equation 4. Shown for comparison with the prediction are a raster plot (**C** and a PST histogram(**E**) of the actual response of the neuron for 10 stimulus presentations. Shading in D and E indicate one standard deviation above and below the mean prediction and the mean PST histogram respectively, obtained with the bootstrap technique. The units of the STRF depicted in B are spikes/sec/Hz/Pa. Blue pixels in the STRF indicate suppression, white pixels indicate no change, and red pixels indicate elevation of discharge rate. Unit *d-46d, class 1*.

FIG. 6. Response of one unit to an anchored tone sequence. Dot rasters (*left*) and schematic spectrograms of each tone in the anchored tone sequence overlayed on the *STRF* of the unit (*right*). Stimuli labeled as *Hx* (where *x* indicates the component number at *BF*) are tones whose partials are in a harmonic sequence while those labeled as *SRx* are tones whose partials are in an inharmonic sequence.Unit *d-76b, class 1*.

FIG. 7. Predictability of temporal patterns of discharge for unit of Figure 6. **A** Actual (*solid*) PST histograms overlayed on predicted (*dotted*) PST histograms for each stimulus in an anchored tone sequence. Dotted red lines indicate stimulus onset and offset. Predictions are normalized by the maximum across stimuli of all prediction waveforms and actual PST histograms are normalized by the maximum across stimuli of all actual responses, because the magnitudes of the predictions are consistently greater than the magnitudes of the PST histograms. **B** Schematic spectrograms for stimuli of the complex-tone sequence, overlayed on the *STRF* of the unit. **C** Distribution of the correlation coefficient (ρ) between predicted and actual PST histograms for harmonic stimuli pooled together and inharmonic stimuli pooled together. Bootstrap resampled estimates of ρ are included to give many more points in the distributions than the number of stimuli of each type.

FIG. 8. Predictability of temporal patterns of discharge in another unit, in the same format as Figure 7. Each waveform (predicted and actual) is normalized by the maximum across all predicted and actual responses, thus preserving relative magnitudes across stimuli. **D** Distribution of distances between harmonic and inharmonic stimuli for all units in population. Unit *d-46d, class 2*

FIG. 9. Distribution of distances between harmonic and inharmonic stimuli for all units in population. Filled square on the abscissa indicates the distance for the unit in Figure 6 and Figure 7 while the filled circle indicates the distance for the unit in Figure 8.
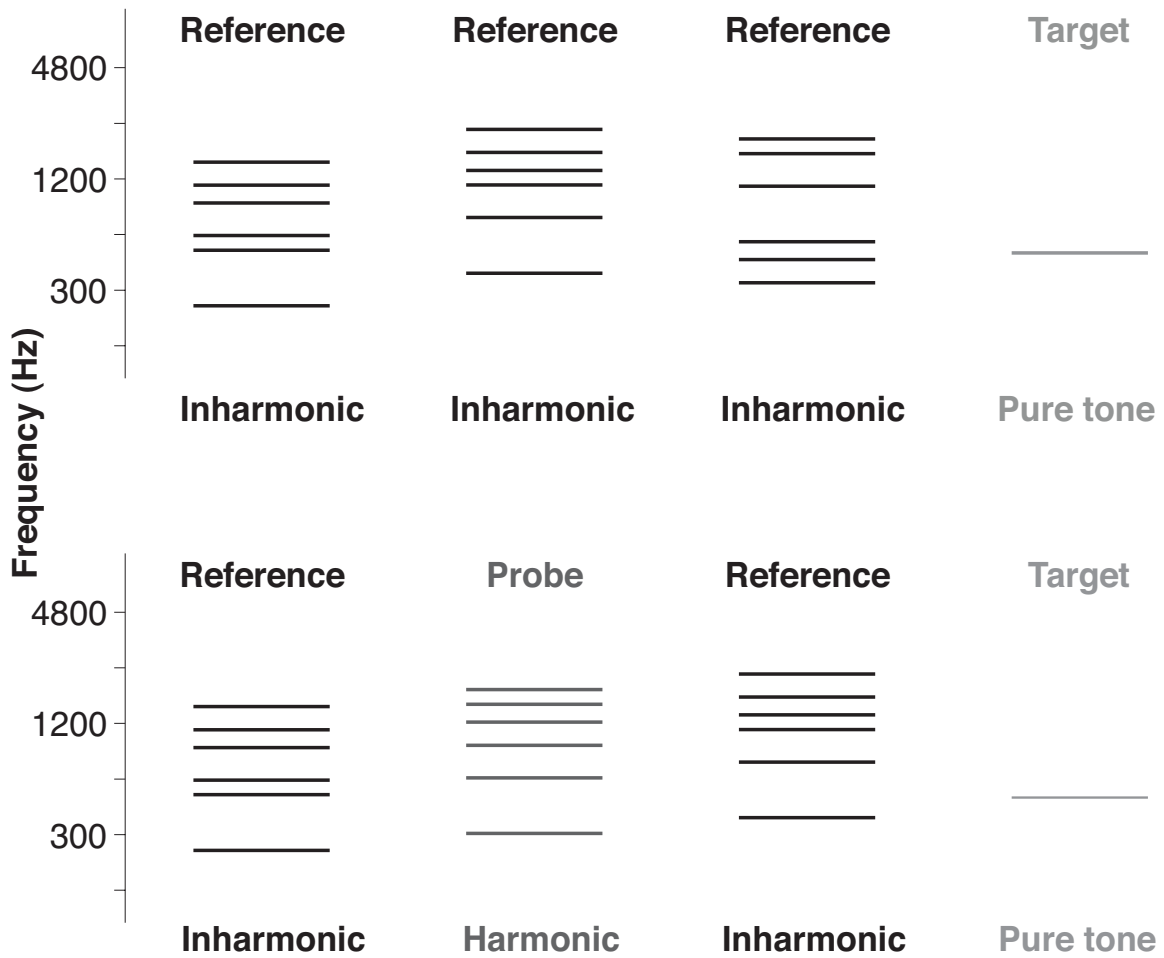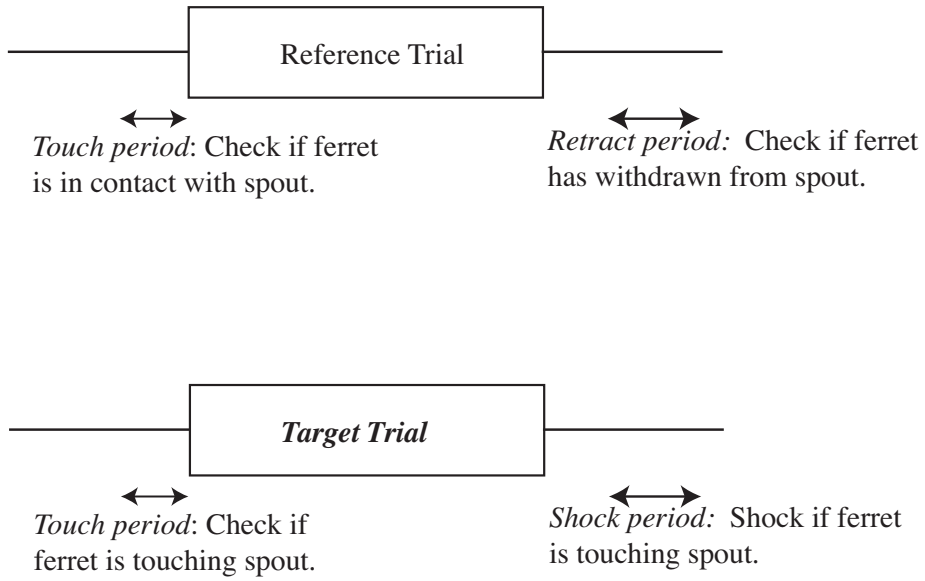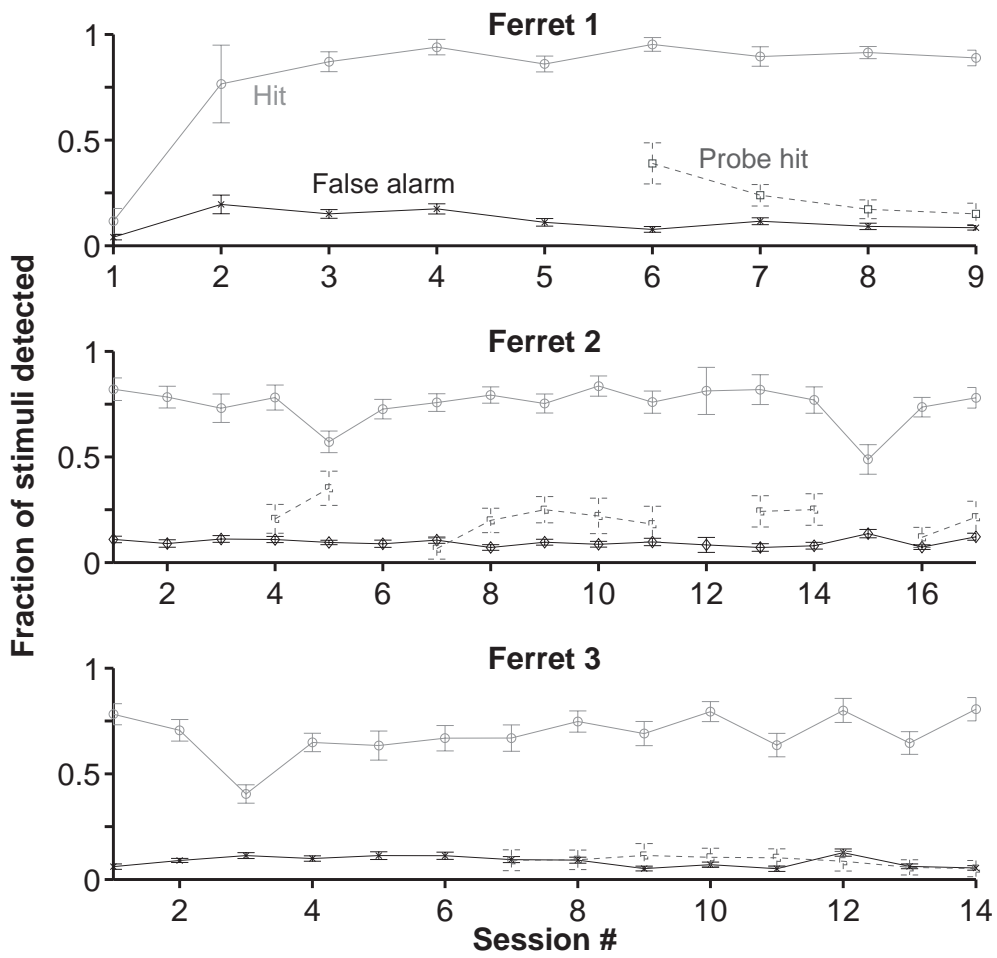
FIGURE 1

A

ITI | Reference | ITI | Reference | ITI | Reference | ITI | *Target* | ITI | - - - - - | ITI | Reference
Pause

B

Reference Trial

Touch period: Check if ferret
is in contact with spout.

Retract period: Check if ferret
has withdrawn from spout.

Target Trial

Touch period: Check if
ferret is touching spout.

Shock period: Shock if ferret
is touching spout.

FIGURE 2

FIGURE 3

FIGURE 4

A Stimulus line spectrogram

B *STRF*

C Actual raster

D Predicted response

E Actual PST hist

Frequency (Hz)

Spikes/second

Time (ms)
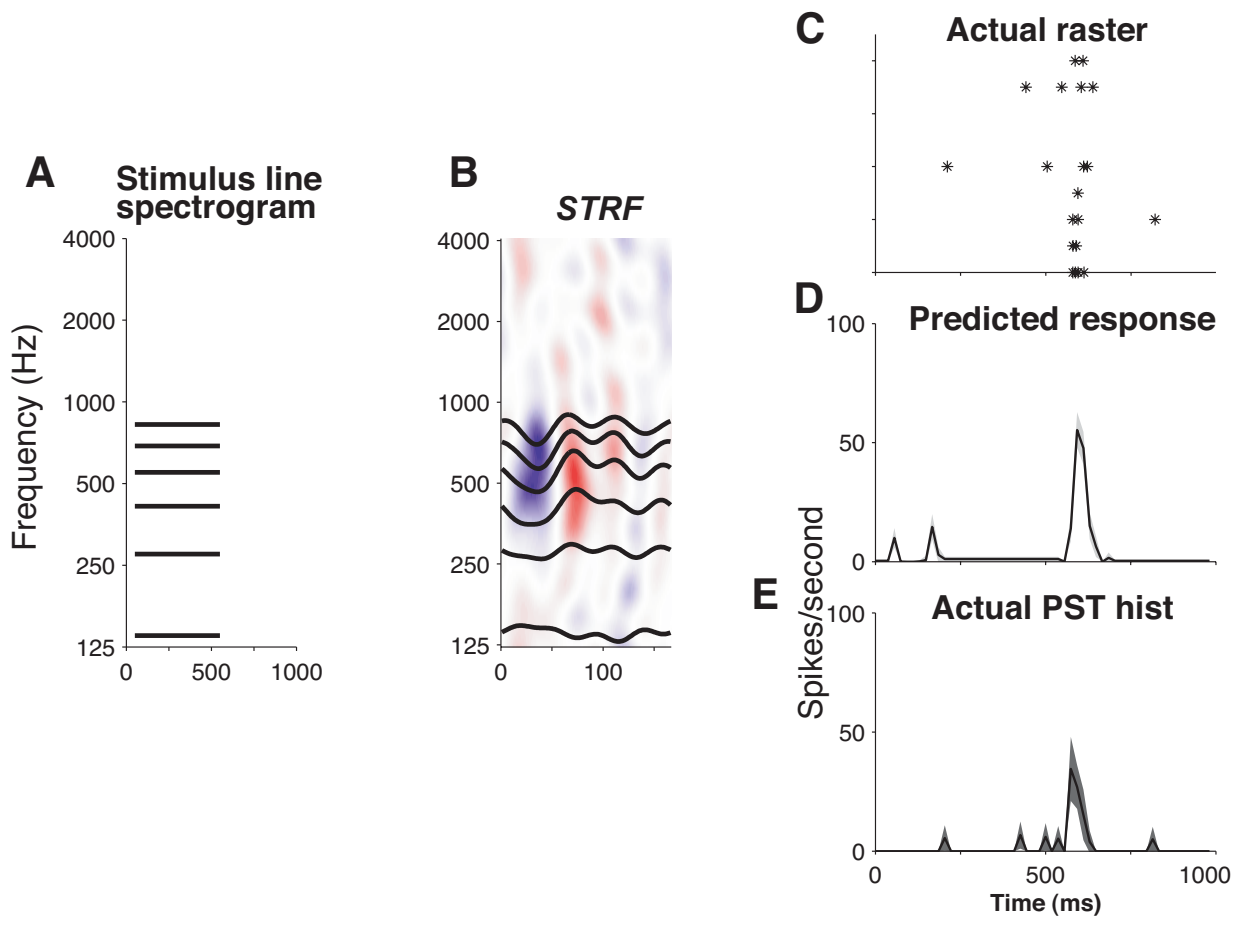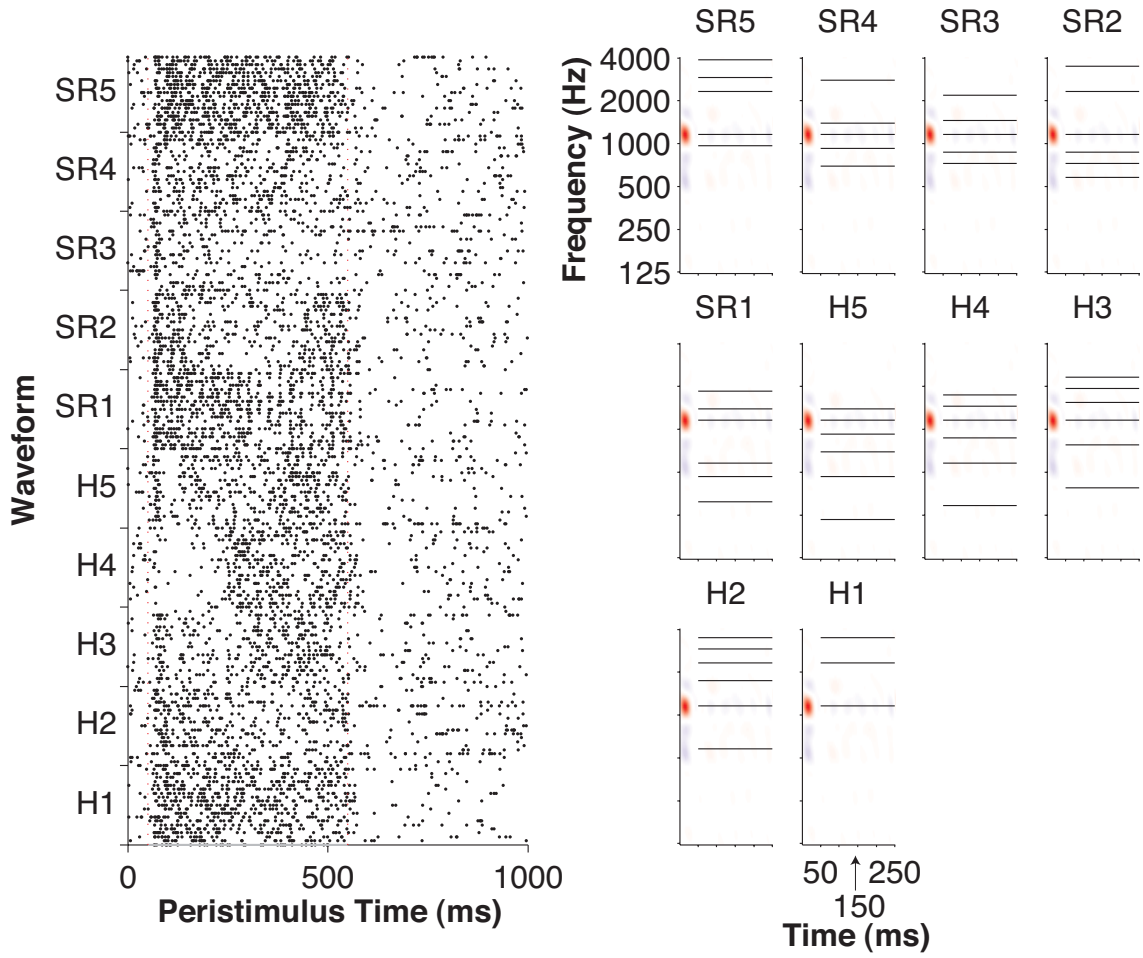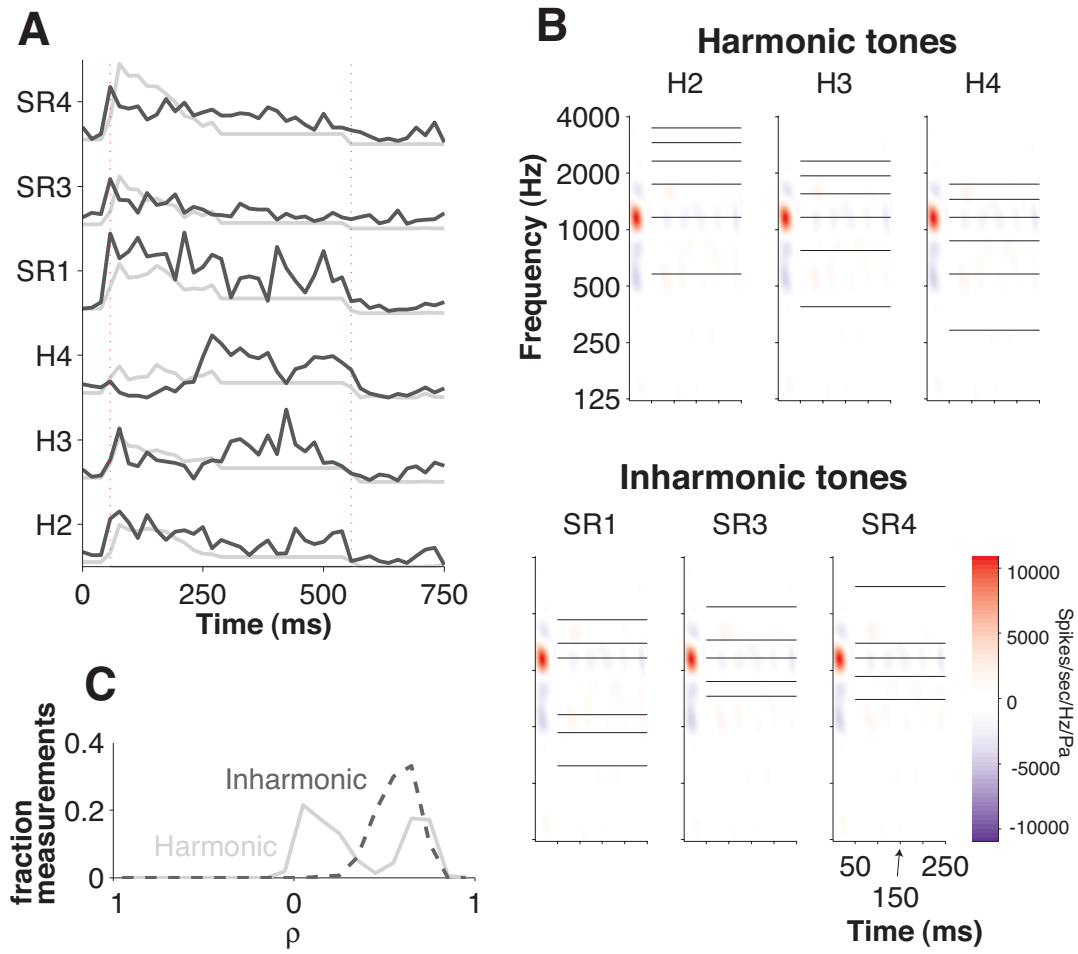
FIGURE 5

FIGURE 6

FIGURE 7

FIGURE 8

FIGURE 9