

Perception and cortical neural coding of harmonic fusion in ferrets

Sridhar Kalluri,^{1,2,*} Didier A. Depireux,² and Shihab A. Shamma¹

¹Institute for Systems Research

University of Maryland

College Park,

MD 20742

²Department of Anatomy and Neurobiology

University of Maryland

20 Penn St - S251

Baltimore,

MD 21201

(Dated: March 18, 2007)

Abstract

Despite much attention, the cortical representation of harmonic sounds has eluded explanation. 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 tones. In baseline sessions, ferrets detected a pure tone terminating a sequence of inharmonic tones. After they reached proficiency in the baseline task, a small fraction of the inharmonic tones were replaced with harmonic tones. Some of the animals confused the harmonic tones with the pure tones at twice the false-alarm rate, indicating that ferrets can fuse the components of a harmonic tone. Experiment 2 sought correlates of harmonic fusion in single neurons of primary auditory cortex (AI) and anterior field (AAF), by contrasting responses to harmonic tones with those to inharmonic tones in the awake alert ferret. The effects of spectrotemporal filtering were accounted for by using the measured spectrotemporal receptive field to predict responses and by seeking correlates of fusion in the predictability of responses. Only twelve percent of units sampled distinguished harmonic tones from inharmonic tones. The paucity of such units suggests that processes mediating harmonic fusion and pitch perception are unlikely to be centered in AI or AAF.

PACS numbers: 43.66.Hg, 43.66.Gf, 43.64.Qh

Keywords: pitch, auditory cortex, neural coding, psychoacoustics

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 (von Helmholtz, 1954). 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 separate auditory objects (Bregman, 1990). An example of this brain function is the improvement obtained in the ability to distinguish two talkers when a fundamental frequency difference is imposed (Brokx and Nootboom, 1980). The highly salient pitch evoked by spectrally resolved 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 (Hartmann *et al.*, 1993; Houtsma and Smurzynski, 1990; Moore and Glasberg, 1985). 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, template matching operates on the neural activity due to the resolved components of the complex tone (e.g., Goldstein (1973); Srulovicz and Goldstein

*Corresponding author, current address: Starkey Hearing Research Center, 2150 Shattuck Ave, Suite 408, Berkeley, CA 94704; Electronic address: sridhar_kalluri@alum.mit.edu

(1983); Terhardt (1974, 1979); Wightman (1973)). Such models predict well that the pitch of harmonic complex tones is not changed by the absence of the fundamental component; this is an important attribute for a model of pitch to possess (de Cheveigne, 2005). There is evidence that a template-matching operation also underlies the perceptual fusion of harmonically related components of a complex tone (Brunstrom and Roberts, 1998, 2000; Lin and Hartmann, 1998).

While spectral fusion and periodicity pitch of complex tones are closely related, previous studies 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). Primary auditory cortex (AI) has failed, on the whole, to yield compelling correlates of periodicity pitch, although a study by Bendor and Wang (2005) in non-primary auditory cortex of marmosets has yielded promising results that remain to be independently replicated and confirmed. See also further reservations discussed in McAlpine (2004).

In recordings from single neurons, Bendor and Wang (2005) found responses in rostral field tuned to the fundamental frequency of harmonic complex tones with and without the missing fundamental component. Such tuning to the fundamental frequency regardless of the physical presence of that component is suggestive of these neurons being tuned to pitch. Nevertheless, it is unclear if these neurons respond differently to harmonic tones than to inharmonic tones, as a pitch-selective or harmonicity-selective neuron might be expected to do.

Studies of AI have failed to yield harmonicity-selective neurons. Recordings from single units of monkey AI in response to missing-fundamental harmonic complex tones failed to show any rate tuning to the fundamental frequency (Schwarz and Tomlinson, 1990). 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 a weak pitch known as residue pitch. Consistent with observations from single-unit recordings (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 *et al.*, 2002; Schulze and Langner, 1997a,b) and of humans (Langner *et al.*, 1997). These findings may indicate the presence of a map of periodicity pitch, but three aspects of the experiments make the assertion inconclusive.

1. The neurons forming the map of envelope periodicity were primarily sensitive to high-frequency energy (> 5 kHz), a frequency region that does not evoke periodicity pitch in human listeners.
2. 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, 1976); Flanagan and Gutman (1960)).
3. Rather than reflecting the periodicity pitch due to the spectrally 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, 2004).

The latter most of these reservations is also applicable to observations in monkey AI of similar responses to a pure tone and a missing-fundamental harmonic complex of the same pitch (Riquimaroux, 1995; Riquimaroux and Kosaki, 1995). Sensitivity to harmonic combinations of resolved components 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 have focused on correlates of periodicity pitch. Several studies have yielded promising candidate neurons that

encode pitch, although these findings need to be further understood and replicated.

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 components 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 feline and primate 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 entity¹. The first experiment in this paper uses a conditioned avoidance behavioral-testing paradigm to specifically test the assumption that ferrets automatically fuse harmonic components of complex tones. Experiment 2 seeks neural correlates of harmonic fusion in ferret auditory cortex, specifically in areas AI and in anterior field (AAF).

II. EXPERIMENT 1: PERCEPTION OF HARMONIC COMPLEX TONES

A. Rationale

In order to determine if ferrets can automatically fuse components 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 components. 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, detecting the probes at a rate greater than the false-alarm rate (i.e., the rate at which references were inadvertently detected as if they were targets) 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.

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-restricted ferret licked water from a continuously dripping spout while listening to reference sounds. At random intervals, an easily distinguishable target sound was presented fol-

lowed by a light shock to the tongue delivered through the reward spout. Such pairings of stimulus and shock helped the ferret learn to break contact with the spout when it heard a target. The continuous water reward encouraged 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 constituted safe trials because it could drink from the spout without getting shocked. On the other hand, target stimuli were warning trials, because they warned the ferret to break contact with the spout.

A computer registered successful withdrawal from the spout following a target as a *hit* and failure to withdraw as a *miss*. Because the animal occasionally withdrew in the absence of a target, a *false-alarm* rate was determined by registering how often it withdrew from the reward spout for reference trials. Both target and reference trials were ignored when the ferret did not contact the reward spout just before the trial; this was done in order to ensure that performance was evaluated only for trials on which the ferret attended to the stimuli. Figure 2 helps understand these measures by illustrating the presentation of trials and the timing of response intervals, while Table I shows which behaviors led to the different response categories.

During testing, two kinds of trial sequences were 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 was randomized such that the probability of the target sound occurring in trial position 2 through 7 was constant. This randomization was done so that the ferret could not preferentially respond on trials occurring at any given position. Sham sequences were 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 were presented in exactly the same way as reference trials by replacing 10% of reference sounds by probe sounds. Responses on the probe trials were scored in the same way as those on the reference trials. *Probe hits* and *probe misses* were equivalent to false alarms and safe responses respectively on reference trials.

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 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 components of harmonic complex tones so that the two types of complex tones elicited comparable percepts of roughness. The lowest component of the inharmonic complex tones was constrained between 150 Hz 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 prevent the use of intensity cues. Stimuli were played at a nominal level of 70 dB SPL; this was the overall level of the complex tones, so individual components were lower in level. Levels were 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 components

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

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. 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 does not hold. First, the parameters of the harmonic complex tones were matched in 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. 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 components 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. 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, gradually learning that the harmonic probes were not associated with the shock, 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 more slowly learnt that harmonic probes were not associated with the 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 become good at the baseline task.

III. EXPERIMENT 2: CORRELATES OF HARMONIC FUSION IN AUDITORY CORTICAL NEURONS

A. Rationale

To seek neural correlates of harmonic fusion, we recorded single-unit activity from primary auditory cortex (AI) and auditory anterior field (AAF) to a sequence of harmonic and inharmonic complex tones where 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 component 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) neural 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.

The effect of a harmonic context on the responses can be confused with the effects of spectrotemporal integration of stimulus spectrum by the spectrotemporal tuning of a cortical neuron (e.g., interaction of inhibitory sidebands with balance of spectrum above and below BF) (Depireux *et al.*, 2001; Kowalski *et al.*, 1996; Schreiner and Calhoun, 1994; Shamma *et al.*, 1995). 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

Experiments were performed at the University of Maryland, College Park and the University of Maryland, Baltimore. Animal preparation and neural-recording procedures were approved by the IACUCs of both institutions. Surgical procedures are described in detail in other publications (Dobbins *et al.*, 2006; Fritz *et al.*, 2003). We give a brief outline of the recording procedure here. During daily recording sessions lasting 3 to 6 hours, ferrets were trained to lie motionless in a

custom apparatus that restrained them. The auditory cortex was accessed through craniotomies ranging in diameter from less than 0.5 mm to 6 mm. We took precautions to maintain sterility of the craniotomies at all times.

- In some experiments, the auditory cortex was accessed through a small hole of diameter less than 0.5-mm. 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 focus on 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.
- In other experiments, we obtained recordings from both AAF and AI. In these experiments, the recordings were through relatively large craniotomies of diameter 5-6 mm. Larger craniotomies permitted us to identify recording locations as being from AI or AAF based on position relative to sulcal landmarks on the surface of the brain (Kowalski *et al.*, 1995). Such attribution of location would not have been possible with smaller craniotomies because sulcal landmarks are not readily visible.

A variety of different methods were used for the recordings. These methods are described below, but they're not distinguished in the reporting of results because they did not lead to noticeably different findings. Most of the recordings in both College Park and in Baltimore were with single parilyne-coated tungsten microelectrodes having resistance ranging from 2 to $7M\Omega$ at 1 kHz. Electrode penetrations were perpendicular to the surface of the cortex. We used a hydraulic microdrive under visual guidance via an operating microscope to position the electrodes. A few recordings were made with NeuroNexus Technologies silicon-substrate linear electrode arrays (4 arrays, with each array having 4 electrodes in a straight line) or silicon-substrate tetrode arrays (4 tetrode arrays, with each array having four closely spaced electrodes in a diamond-shaped configuration). Recordings yielded spikes from 1 – 3 neurons on a single electrode that were sorted offline with a combination of automatic spike-sorting algorithms (Lewicki, 1994; MClust-3.3, A.

D. Redish et al; KlustaKwik, K. Harris) and manual techniques. A spike class was included as a single unit for further analysis if less than 5% of interspike intervals were smaller than 1 ms, the putative absolute refractory period, and the spike waveform had a bipolar shape.

In some experiments, 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* for every recording session. An automatic calibration procedure gave flat frequency responses below 20000 Hz. In other experiments, sounds were delivered by a speaker (Manger) placed 1 meter from the head of the ferret. Stimuli were generated by a computer and sound levels were calibrated with a 1/2"-pressure microphone (ACO Pacific) at the position of the ferret's head. Apart from a handful of cases, most of the recordings in these latter experiments were obtained without the sound system having been equalized to a flat frequency response, because levels were fairly flat even when not equalized.

2. Stimuli and analysis

a. 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 an estimate of its *BF*. The discharge rate was measured as a function of the level of the *BF* pure tones 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 neuron's spectrotemporal filtering characteristics was estimated at the same level as the anchored-tone sequence by measuring the *STRF* with temporally-orthogonal-ripple-combination (*TORC*) stimuli (Klein *et al.*, 2000), which is described below.

b. 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 phase were part of the sequence, where each tone had a different component number at the anchor frequency. For most units, the first through sixth components were at the anchor frequency, but a few units had only the second through fifth or first through fifth components 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 components 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 components 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 components 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.

In our later recording sessions, we imposed a slow sinusoidal amplitude modulation (4 - 20 Hz) with a modulation depth of 0.5 upon the complex tones. All components were modulated coherently. We imposed this modulation to ensure that neurons responded vigorously throughout the complex tone, because cortical neurons are often unresponsive during the sustained portions of static stimuli (Clarey *et al.*, 1992; Shamma *et al.*, 1995) as was indeed the case for the majority of single units recorded during our early recording sessions. In contrast, most units responded vigorously to the amplitude-modulated complex tones.

All components of the harmonic and inharmonic complex were presented at the same level.

c. Characterizing linear processing of spectrotemporal envelopes with STRFs Underlying the measurement of a *STRF* is the observation (Kowalski *et al.*, 1996; Schreiner and Calhoun, 1994; Shamma *et al.*, 1995) that responses of many 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

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

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 neural response 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* (Depireux *et al.*, 2001; Kowalski *et al.*, 1996), 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 periodicity and ripple density) whose two-dimensional inverse Fourier transform is the $STRF$. $TORC$ s are special superpositions of moving ripples such that two components having different ripple densities do not share the same ripple periodicity. This special combination of moving ripples enables rapid measurement of the $STRF$. We used random-phase tones densely spaced on a log-frequency axis as carriers (100 tones/octave, spanning 5 octaves in the early recordings and spanning 7 octaves in later recordings) 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 steady-state. Consequently, the $STRF$ quantifies changes of the discharge rate above and below a steady-state rate.

2. The mean of the *STRF*s is set to zero because the *STRF* is designed to predict response modulations, not the steady-state response. In order to predict the steady-state discharge rate, it is necessary to measure the terms of the *STRF* at a ripple periodicity of 0 Hz; 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 0-Hz ripple periodicity terms can be estimated separately using static ripples presented at a range of phases. We estimated such terms of the *STRF* for 13 units in some of the later experiments. However, because results from these units did not differ from other units, we do not distinguish them in the results.

d. 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 that was not explainable by spectrotemporal filtering. 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, we obtained $p_{strf}[n]$, which is simply the convolution of the stimulus spectrogram with the *STRF*, in two steps.

1. Convolve stimulus envelope $e[n]$ (flat in the early experiments that used unmodulated complex tones and sinusoidal in the later experiments that used modulated complex tones, both having 10-ms cosine-squared onset and offset ramps) with horizontal slices of the *STRF* at the frequencies x^i of the components, $STRF[n, x^i]$, to get the contribution of individual

components to the overall prediction.

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

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

2. Combine contributions of individual components to get the overall prediction from the $STRF$ as the mean of the individual contributions.

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

Finally, the predicted PST histogram was

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

where $g(\cdot)$ indicates half-wave rectification. Because the $STRF$ does not predict the steady-state discharge rate (except in the handful of measurements where we did measure responses to static ripples), we used the measured steady-state discharge rate r_{ss} instead of p_{ss} . In Figure 5E, 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 D 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.

We used the cross-covariance between the response $r[n]$ and the prediction $p[n]$ as a measure of the quality of the prediction. We defined it to be

$$\rho[m] = \frac{cov(r[n+m] \cdot p[n])}{\sigma_r \cdot \sigma_p}$$

Because the covariance in the numerator is normalized by the standard deviations, $\rho[m]$ is insensitive to the magnitudes of $r[n]$ and $p[n]$. We use $\rho = \max(\rho[m])$ to summarize the relationship of prediction and response independent of temporal misalignments between the waveforms.

To quantify the extent to which a neuron distinguished a harmonic tone from each of the inharmonic stimuli, we used a measure analogous to the *d-prime* statistic from signal detection theory. We defined this measure to be the difference between the means of the distributions of

ρ for each of the stimuli normalized by the geometric means of the standard deviations of the distributions, with the distributions having been generated by bootstrap replication.

$$d'_\rho(Hx, IHx) = \frac{\overline{\rho(Hx)} - \overline{\rho(IHx)}}{\sqrt{\sigma_{\rho(Hx)} \cdot \sigma_{\rho(IHx)}}}$$

We did not compute this quantity when either stimulus of a Hx, IHx pair (denoting harmonic and inharmonic tones respectively) did not evoke a visible change in discharge rate or a visible phase-locked response. Figure 7C gives an example of distributions of ρ generated by bootstrap resampling for a harmonic tone and several inharmonic tones. We take $d' = 1$ as the threshold at which the two distributions are significantly different (equivalent to 79% correct identification in a 2-alternative forced-choice discrimination) to determine if a neuron reliably distinguished a harmonic tone from all of the inharmonic stimuli. We revisit this example in the Results section below.

C. Results: A small fraction of cortical neurons distinguished harmonic complex tones from inharmonic complex tones

We report data from 86 units in 9 ferrets. Their BF s ranged from 210 Hz to 13600 Hz, with 11 having BF greater than 5000 Hz, 20 having BF greater than 3000 Hz, and 36 having BF greater than 1500 Hz. Of these units, 51 were from AI and 22 were from AAF. Thirteen units were either in AI or AAF; we were unable to identify the location.

Figure 6 shows the responses of a single unit to unmodulated 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 'IH'. 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 components at the AF 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 IH3 and IH4 than to any of the other stimuli.

Interestingly for this unit, the response to stimulus IH1 shows temporal structure in the discharge pattern. This temporal structure reflects synchronization of discharges to envelope mod-

ulations 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, where the fundamentals exceeded 250 Hz, 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 via synchronization to the envelope; 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 (Elhilali *et al.*, 2004; Liang *et al.*, 2001; Schulze *et al.*, 2002; Steinschneider *et al.*, 1994). No other unit in our population exhibited such discharge patterns.

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 overlaid on predictions for the unit of Figure 6, with line spectrograms of the stimuli shown overlaid on the *STRF* of the unit in B. The response was better predicted for the inharmonic stimuli than for stimuli H3 and H4 because the *STRF* was unable to predict the late response to these harmonic stimuli. We summarized the linear predictability of the response using the maximum cross-covariance ρ (see Methods), between the response PST histogram and the predicted waveform. Figure 7C shows distributions of ρ for stimulus H4 and for all the inharmonic stimuli. We quantified the extent to which this neuron distinguished H4 from each of the inharmonic stimuli using d'_ρ (see Methods). If we take $d' = 1$ to be the threshold at which two distributions are significantly different, then this neuron reliably distinguished H4 from all of the inharmonic stimuli. Figure 7D summarizes d'_ρ for each of the harmonic stimuli compared to all of the inharmonic stimuli. The minimum d'_ρ for a given harmonic tone is a convenient measure of how well the neuron discriminated the harmonic tone from the inharmonic tone having the most similar response. This measure is greatest for H4 of all the harmonic tones, which means that this neuron was best at discriminating H4 from inharmonic tones.

Figure 8 shows an example of responses and predictions from a unit for amplitude-modulated complex-tone stimuli. Figure 8A shows period histograms of responses to a harmonic (H3) and

an inharmonic tone² along with a single period of predictions for the same stimuli. Both the magnitudes and phases of the responses are predicted well by spectrotemporal filtering of stimuli by the STRF of this neuron. Figure 8B shows histograms of bootstrap-replicate ρ for stimulus H3 and the inharmonic tone of panel A. The distributions of ρ for these stimuli overlap a great deal, suggesting that the predictability of temporal discharge pattern does not readily distinguish H3 from the inharmonic tone. Figure 8C summarizes d'_ρ for each of the harmonic tones compared to all of the inharmonic tones. The minimum d'_ρ for each harmonic tone is less than 1, confirming the observation for stimulus H3 that this neuron was not effective at discriminating any of the harmonic tones from inharmonic tones.

Most units in the population were very ineffective at discriminating the harmonic tones from the inharmonic tones; that is, they were mostly like the unit of Figure 8 rather than the unit of Figure 7. Figure 9 summarizes for our entire population of neurons how well the predictability of the temporal pattern of discharge discriminated the harmonic tones from the inharmonic tones. Histograms of minimum d'_ρ in Figure 9A show that values exceeding 1 were infrequently observed, indicating that few neurons discriminated any of the harmonic tones from inharmonic tones. Figure 9B gives another view of the same information. The figure shows the minimum d'_ρ for each harmonic tone as a function of the *BF* of the unit, while identifying the harmonic tone by digits corresponding to the number of the anchor component. The 13 units that had a minimum d'_ρ greater than 1 had *BF*s between 500 Hz and 10000 Hz (all but one unit having *BF*s less than 4400 Hz). Location of the units (signified by color of the symbol) that discriminated harmonic tones from inharmonic tones could be either AI or AAF. In summary, these few discriminative units weren't biased to a particular *BF* range nor were they likely to occur in a particular cortical location.

The maximum cross-covariance quantifies the similarity of the shapes of the predicted waveforms and the PST histograms. However, its quantification leaves out how the magnitude of the predicted waveform compares to the magnitude of the time-varying component of PST histograms, or any differences in latencies of the prediction and actual response. For modulated complex tones it fails to quantify synchronization of response to the amplitude modulation. Moreover, the steady-

state discharge rate is not predicted by the *STRF*, but it must be affected by the interaction of the stimulus spectrum with the frequency tuning of the neuron. We accounted for these various other properties of response predictability in a manner similar to our examination of the maximum cross-covariance (not shown). Our conclusions based on the maximum cross-covariance were not significantly altered – between 10% and 15% of the units were capable of discriminating one of the harmonic tones from the inharmonic tones with these other measures.

D. Discussion

About 10% – 15% of units in AI and AAF responded differently to harmonic tones and inharmonic tones. Given how readily humans (von Helmholtz, 1954) and ferrets (Experiment 1) perceive harmonic sounds differently than inharmonic sounds, it seems unlikely that such a small fraction of neurons mediates such a strong percept. The failure to find a large percentage of neurons that discriminate harmonic sounds from inharmonic sounds leads us to conclude, consistent with the conclusion of previous studies, that AI neurons do not compute harmonicity. The correspondence of AAF in ferrets to cortical fields in other animals is less certain (Bizley *et al.*, 2005). If AAF corresponds to the rostral field in primates, then our failure to find a significant percentage of harmonicity-selective neurons may be surprising given the promising observations of neurons tuned to the fundamental frequency of harmonic tones in marmosets (Bendor and Wang, 2005).

There are several possible reasons for the lack of congruence between the marmoset data and the ferret data. Given that Bendor and Wang (2005) found neurons tuned to the fundamental frequency of harmonic complex tones in rostral field of marmosets, one might be surprised that we found few neurons that were harmonicity-selective in AAF of ferrets. One possible reason for the difference is that the marmoset neurons were not harmonic selective but instead were selective to waveform modulation rate. Other reasons might be that AAF in ferrets does not correspond to rostral field in marmosets or that the marmoset data were mostly at very low frequencies (< 0.8 kHz), while our data are at higher frequencies (mostly between 1 and 5 kHz – see Figure 9B). Future ferret studies having more data from the low-frequency region of AAF may yield harmonicity-

selective neurons.

One previous study (Fishman *et al.*, 2000) used stimuli that resembled the harmonic complex tones in the anchored tone sequence. Fishman *et al.* 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).

In general, the present study differs in outlook from most previous studies of the cortical neural coding of harmonicity. Previous studies have sought correlates of harmonicity by focusing on two features of the percept: 1) periodicity pitch is at the common fundamental of a sequence of harmonics, and 2) the pitch corresponds to that of a pure tone at the fundamental frequency. The former property is expected to endow the putative pitch neuron with having inputs across a broad range of frequencies, while the latter property is expected to endow the neuron's tuning to pure tones to be at low frequencies below 1100 Hz (the range of periodicity pitch in humans). The present study exploits the fact that the brain uses harmonicity as a cue to group individual spectral components of a sound. It assumes that the fusion of harmonic components leaves an imprint on the neural representation of the individual components themselves; i.e., the brain maintains a pitch- or harmonicity-tagged representation of sound that is a by-product of the harmonic fusion process. Neurons that have an imprint of this harmonicity-tagging span a broad frequency range of best tuning up to 5000 Hz (the existence region for periodicity pitch (Ritsma, 1962)). This frequency range is greater than the expected frequency range of tuning for neurons whose selectivity to fundamental frequency corresponds to the pure-tone best frequency. The larger candidate pool of

neurons in the *fusion-based method* compared to the *fundamental frequency tuning equals pure-tone tuning method* can be seen as an advantage when seeking correlates of harmonicity. On the other hand, a nice feature of the latter method is that it can better reveal the actual neurons that combine harmonic components across frequency.

IV. GENERAL DISCUSSION

Our results show that ferrets hear harmonic complex tones as fused, unitary entities. However, this fusion does not leave a compelling imprint upon single neurons in AI or AAF. The absence of pervasive correlates of harmonic fusion in our data is consistent with the failure of previous studies to find conclusive evidence for pitch-tuned neurons in AI.

Several single-unit physiology studies have failed to find neurons that are selective to periodicity pitch in AI (Schwarz and Tomlinson, 1990; Steinschneider *et al.*, 1994). Recent neuromagnetic and functional magnetic resonance imaging studies also support the conclusion that periodicity pitch is not extracted in AI. Instead, investigations with regular-interval noise and filtered harmonic complex tones suggest that pitch is extracted at the anterolateral border of AI in humans (Krumbholz *et al.*, 2003; Patterson *et al.*, 2003; Penagos *et al.*, 2003; Warren *et al.*, 2003). However, this conclusion about the role of the anterolateral border of AI has recently come under doubt (Hall and Plack, 2006, 2007).

The neurons that confer a special perceptual status to harmonic sounds remain to be conclusively identified. Emerging functional imaging studies may provide some guidance on the cortical areas that should be investigated in future single-unit studies.

APPENDIX: 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

$$S(t, x) = a_0 + a \cos[2\pi(\omega t + \Omega x) + \psi] \tag{A.1}$$

At each frequency location, the function describes a sinusoidal modulation of sound level 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 system³ (Klein *et al.*, 2000, 2006).

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 periodicities, ω) can evoke overlapping response components that cannot be disambiguated; 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 periodicity $|\omega|$, 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 periodicity contribute only to a single $[\omega, \Omega]$ component of the *STRF*.

ACKNOWLEDGMENTS

SK was supported by T32 DC-00046 and F32 DC-05281, and DAD was supported by RO1 DC-005937 from the National Institute of Deafness and Other Communication Disorders. The project was made possible by a grant from the Bresser Research Foundation to DAD. We thank Shantanu Ray for expert programming and electronics assistance, Yadong Ji, Tamar Vardi, and

Sarah Newman for help with training animals, as well as Heather Dobbins and Barak Shechter for assistance with physiological recordings.

ENDNOTES

1. Spectral fusion of components and their contribution to periodicity pitch are not entirely congruent in human listeners as well (Brunstrom and Roberts, 1998). For example, a component must be mistuned from a harmonic relation by 1.5% to 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 (Ciocca and Darwin, 1993; Darwin and Ciocca, 1992; Moore *et al.*, 1986; Moore and Kitzes, 1985).
2. The period histogram accumulates the number of spikes occurring at a given phase of the modulation period of the amplitude-modulated complex tones. We accumulate all spikes occurring between 100-ms after stimulus onset and the end of the stimulus.
3. For the cortical neurons investigated in this paper, ripple periodicities were less than 32 Hz and ripple densities were less than 1.4 cycles/octave.

REFERENCES

- Bendor, D. A. and Wang, X. (2005). “The neuronal representation of pitch in primate auditory cortex”, *Nature* **435**, 341–346.
- Bizley, J. K., Nodal, F. R., Nelken, I., and King, A. J. (2005). “Functional organization of ferret auditory cortex”, *Cerebral Cortex* **15**, 1637–1653.
- 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.
- Clarey, J., Barone, P., and Imig, T. (1992). "Physiology of thalamus and cortex", in *The Mammalian Auditory Pathway: Neurophysiology*, edited by D. Webster, A. Popper, and R. Fay, 232–334 (Springer-Verlag, New York).
- 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.
- 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, volume 3, 479–583 (Springer Verlag, Berlin).
- 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.
- de Cheveigne, A. (2005). "Pitch perception models", in *Pitch: Neural Coding and Perception*, edited by C. J. Plack, A. J. Oxenham, R. R. Fay, and A. N. Popper, Springer Handbook of Auditory Research (Springer, New York).
- 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.
- Dobbins, H. D., Marvit, P., Ji, Y., and Depireux, D. A. (2006). "Chronically recording with a multi-electrode array device in the auditory cortex of an awake ferret", *J. Neurosci. Methods* **Epub ahead of print**.
- Efron, B. and Tibshirani, R. (1993). *An Introduction to the Bootstrap* (Chapman and Hall, New York).
- 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.

Hall, D. A. and Plack, C. J. (2006). “Searching for a pitch centre in human auditory cortex”, *International Symposium on Hearing* .

Hall, D. A. and Plack, C. J. (2007). “The human ‘pitch center’ responds differently to iterated noise and huggins pitch”, *Neuroreport* **In Press**.

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 *Methods in Comparative Psychoacoustics*, edited by G. Klump (Springer-Verlag Telos).

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., Simon, J. Z., Depireux, D. A., and Shamma, S. A. (2006). “Stimulus-invariant processing and spectrotemporal reverse correlation in primary auditory cortex”, *J. Comput. Neurosci.*

20, 111–136.

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.

Kowalski, N., Vernsel, H., and Shamma, S. A. (1995). “Comparison of responses in the anterior and primary auditory fields of the ferret cortex”, *J. Neurophys.* **73**, 1513–23.

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., 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.

McAlpine, D. (2004). “Neural sensitivity to periodicity in the inferior colliculus: evidence for the role of cochlear distortions”, *J. Neurophysiol.* **92**, 1295–1311.

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.

Riquimaroux, H. (1995). “How to enjoy music by the brain not by the ears: Neuronal basis of pitch extraction”, *RIKEN Review – Brain and information science* **9**.

Riquimaroux, H. and Kosaki, H. (1995). “Involvement of neurons in the primary auditory cortex in pitch extraction”, *Abstr. Assoc. Res. Otolaryngol.* **18**.

Ritsma, R. (1962). “Existence region of the tonal residue i”, *J. Acoust. Soc. Am.* **34**, 1224–1229.

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**, 651–663.

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. 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.

Srulovicz, P. and Goldstein, J. (1983). “A central spectrum model: a synthesis of auditory-nerve timing and place cues in monaural communication of frequency spectrum”, *J. Acoust. Soc. Am.* **73**, 1266–1276.

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.

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. (1954). *On the sensations of tone as a physiological basis for the theory of music* (Dover Publications, Inc, New York).

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.

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

| 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.

LIST OF FIGURES

- FIG. 1 Stimulus protocols for testing if ferrets automatically fused components of harmonic complex tones. (*Top*) In baseline sessions, ferrets were 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 were i) differences in the quality of perceptual fusion and ii) differences in timbre. (*Bottom*) In probe sessions conducted after ferrets attained proficiency in baseline sessions, a small fraction of the reference sounds were replaced by harmonic-complex-tone probe sounds. If harmonic complex tones were perceptually fused, then they might have occasionally been confused with the pure-tone targets, thereby indicating that ferrets can automatically fuse harmonic components. 38

FIG. 2 **A** Schematic representation of a trial sequence, where the target sound was presented on trial 4. The presentation of trials was paused after a target trial until the ferret returned 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 determined the response class for the trial. If the ferret was not in contact with the spout during the first interval (the *touch* period), the response was classified as an “early” withdrawal and not counted toward computing overall performance on the experiment. If the ferret failed to contact the spout during the second interval (the *retract* period), the trial performance was scored as a “false alarm”. **C** Schematic representation of a target trial shows the same two intervals of time. The first interval played the same role as on a reference trial. The second interval (the *shock* period) was a period when contact with the spout elicited a shock, and the trial performance was scored as a “miss”. The *touch* period was 0.2–seconds long while the *retract* and *shock* periods were 0.4–seconds long in the experiment. The interval between the end of the stimulus and the start of the *retract* and *shock* periods was 0.2–seconds long. Based on a figure from Heffner and Heffner (1995). 39

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) had probe hit rates that were significantly greater than the false-alarm rate during the probe sessions, indicating that they automatically heard harmonic complex tones as being different from inharmonic complex tones. 40

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 BF of the neuron under investigation. If cortical (or upstream) neurons treat harmonic sounds in a special manner, then the component at BF should elicit a qualitatively different response when presented in a harmonic context compared to an inharmonic context. 41

FIG. 5 Predicting responses to complex tones using *STRFs*. The predicted response (**E**) is the convolution of the spectrotemporal envelope of the stimulus (**A**, schematic spectrogram for complex-tone having 6 components) 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 component; ii. Average of the results of the convolutions. These steps form part of an implementation of Equation 4. Shown for comparison with the prediction are a raster plot (**C**) and a PST histogram(**D**) 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/Pa. White pixels in the STRF indicate suppression, gray pixels indicate no change, and black pixels indicate elevation of discharge rate. Unit *d-46d, class 1*. 42

FIG. 6 Response of a unit to an anchored-tone sequence. Dot rasters (*left*) and schematic spectrograms of each tone in the anchored tone sequence overlaid on the *STRF* of the unit (*right*). Stimuli labeled as *Hx* (where *x* indicates the component number at *BF*) are tones whose components are in a harmonic sequence while those labeled as *IHx* are tones whose components form an inharmonic sequence. Unit *d-76b, class 1*. 43

FIG. 7 Predictability of temporal patterns of discharge for unit of Figure 6. **A** Actual (*solid*) PST histograms overlaid on predicted (*dotted*) PST histograms for some select stimuli in the anchored-tone sequence. Dashed vertical gray 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, overlaid on the *STRF* of the unit. **C** Histogram of the correlation coefficient (ρ) between predicted and actual PST histograms for harmonic stimulus H4 (*solid*) and histograms of ρ for each of the inharmonic stimuli (*dashed*), where histograms are obtained from 128 bootstrap replicates of predicted and actual responses. Dotted vertical lines indicate means of the distributions of ρ for stimulus H4 and for the inharmonic stimulus whose response predictability was closest to that of H4. d'_ρ is the difference between such means normalized by the geometric mean of their standard deviations. **D** d'_ρ for each of the harmonic tones relative to each of the inharmonic tones. The minimum d'_ρ for a given harmonic number (*solid circles*) is a measure of how discriminable that particular harmonic tone is from the most similar inharmonic tone in terms of the predictability of the temporal discharge pattern. 44

FIG. 8 Predictability of temporal patterns of discharge in a unit driven by modulated complex tones. **A** A single period of predicted (*dashed*) and actual (*solid*) responses for harmonic tone H3 (*thick gray*) and an inharmonic tone (*thin black*). **B** Histograms of correlation coefficients between 128 bootstrap replicates of predicted and actual responses from panel **A**. d'_ρ for this pair of stimuli is the difference in the means of the histograms normalized by the geometric means of their standard deviations. **C** d'_ρ for each of the harmonic tones relative to each of the inharmonic tones, with *solid circles* indicating the smallest d'_ρ for a given harmonic tone. . . 45

FIG. 9 Summary of how distinguishable harmonic tones are from inharmonic tones across all recordings. **A** Histograms of the minimum d'_ρ for each harmonic tone across all recordings. **B** Minimum d'_ρ for each harmonic tone as a function of the best frequency for the recording. The recording location is indicated by font style (bold: AI, italic: AAF, plain: undetermined) and the identity of the harmonic tone is indicated by the numeric symbol. [COLOR ONLINE] 46

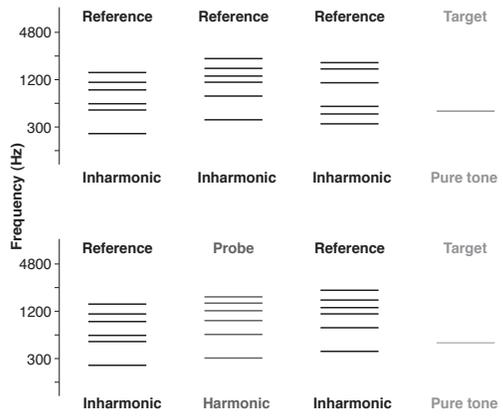


FIG. 1.

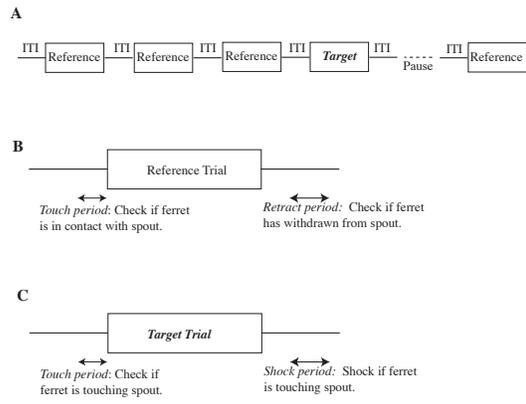


FIG. 2.

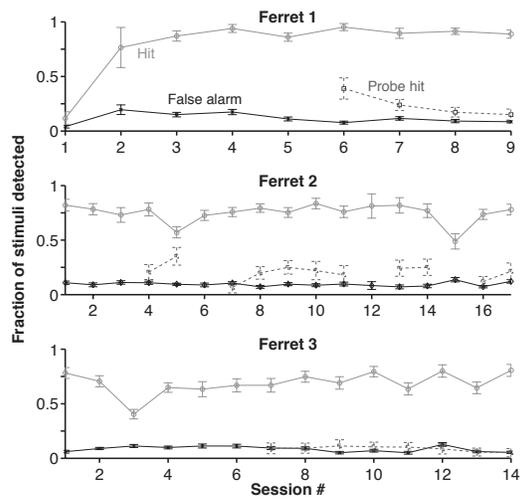


FIG. 3.

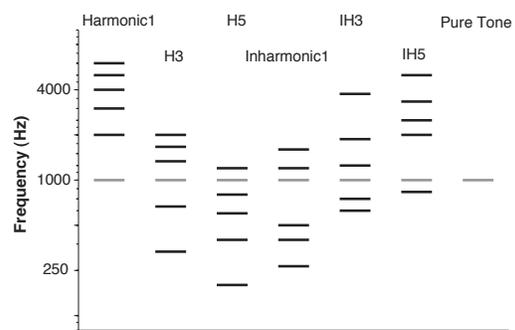


FIG. 4.

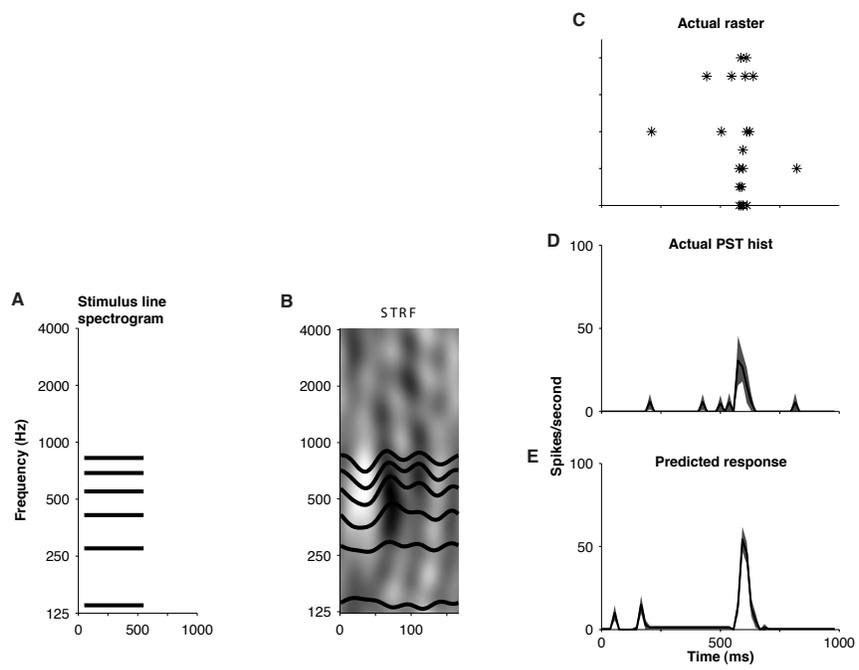


FIG. 5.

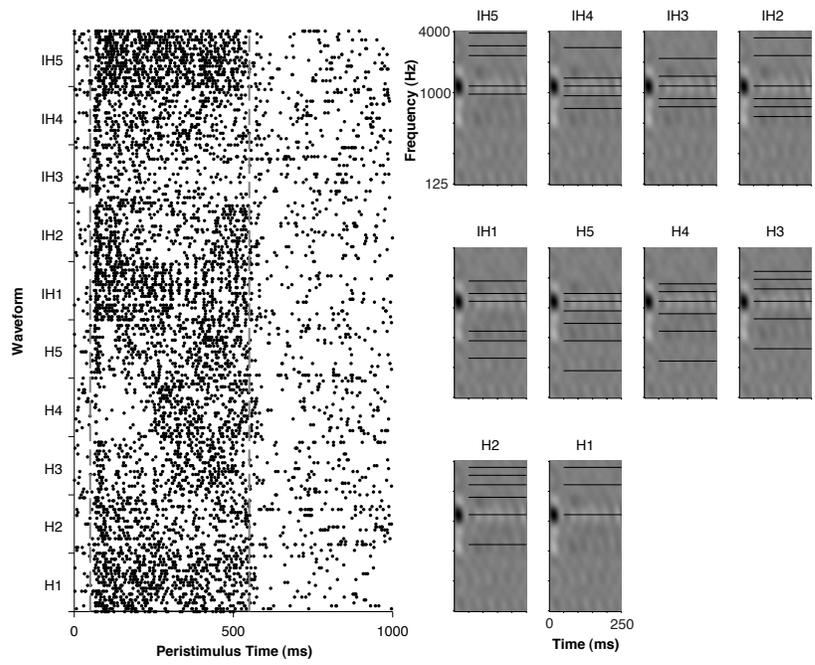


FIG. 6.

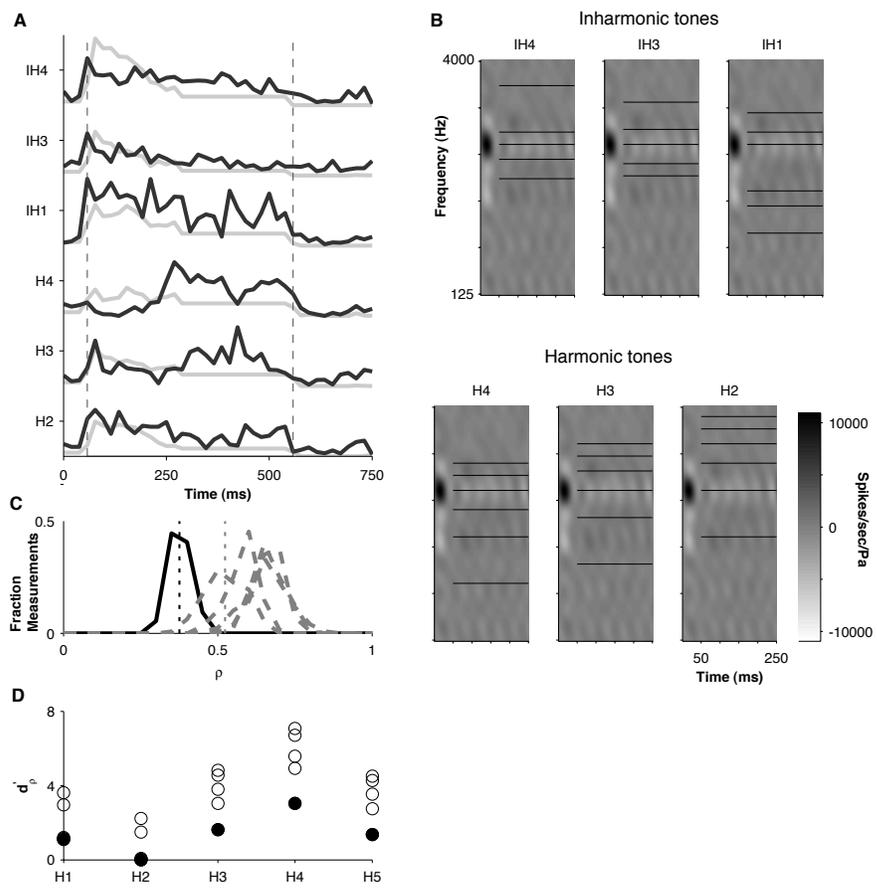


FIG. 7.

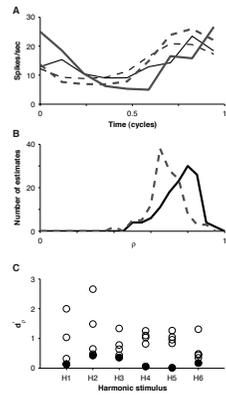


FIG. 8.

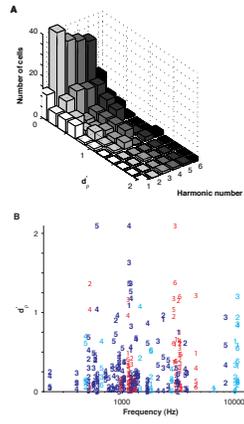


FIG. 9.