Responses to Harmonic and Mistuned Complexes in the Awake Marmoset Inferior Colliculus

by
Kevin Kostlan

A thesis submitted to Johns Hopkins University in conformity with the requirements for the degree of Master of Science in Engineering

Baltimore, Maryland
September 2014

Abstract

The auditory system parses complex acoustic scenes. Often, multiple different sounds overlap in both the spectral and temporal representation at the auditory nerve (AN) level. To separate sources in these cases, the AN information is "regrouped" such that each sound becomes a single percept (Licklider, 1954). The system takes advantage of harmonicity in each source (Scheffers and Maria, 1983: p134) in doing so. Sensitivity to periodicity/harmonicity was found in Bendor and Wang (2010) and Feng (2013) in the awake Marmoset cortex. We used the same preparation in the Inferior Colliculus (IC), which is two levels "below" the cortex, but found no preference for harmonics (n=22). Most units responded whenever there was energy in their receptive field and many were *inhibited* by multi-component complexes such as harmonic stimuli. The rare examples (n=2) that had a strong preference to harmonics over tones did *not* have the sharp tuning properties as did harmonic units in Feng (2013). Furthermore, units were not much more sharply tuned to harmonics than was predicted by a superposition of pure tone responses combined with side-band inhibition and saturation. Although harmonically selective units could be *built* from IC units (that were reasonably sharply tuned) by taking the weighted sums of a *pseudopopulation* (as in May et al., 1998 and Cai et al, 2009) generated from the unit and applying a threshold, the same was also true with a simulated auditory nerve (AN) model from (Zilany et al, 2009). The lack of non-trivial IC responses to harmonics and the success of the AN pseudopopulation method construct to construct harmonically selective units suggests that such selectivity originates *de novo* in the cortex that the IC does not necessarily play a major role in sound regrouping.

Table of Contents

Introduction....1
Dedicated processing....1
Perception of pitch and harmonics....7
Auditory nerve responses to harmonic sounds....9
Origen of pitch/harmonic-sensitive cortical units....13
Motivation....25
Methods....26
Experimental preparation....26
Sound generation....28
Stimuli design....28
Exploratory study and final stimuli design....29
Best frequency, threshold, and map collection....33
Data Analysis....36
Firing rate and “spont”....36
Measures of tuning width and selectivity....39
Concatenating maps....42
Analyzing data on a per-order basis....44
Linear additive model....48
Building a harmonic template unit out of pseudopopulations....50
Cat Auditory-nerve model....54
Results....57

Tuning properties..... 57
Responses of ICC units to tones, harmonics, and mistuned complexes.....59
Harmonic template constructs 73
Discussion/Conclusion.....84
No harmonic selectivity was observed in the ICC.....84
The mysterious world of the ICX.....85
Some high frequency ICC and AN neurons can be used to build harmonic template neurons..... 86
Difficulties with temporal sensitivity due to tuning.....88
High-order non-trivialities..... 90
Multiple sounds.....91
References.....92
Biography and Curriculum Vitae 96

Introduction

It is essential for most animals, including humans, to make inferences about very complex environments. Sophisticated sensory organs evolved to gather the needed information, but extracting meaningful results is still difficult. How does the brain do that? The auditory system is a highly developed and specialized platform with which to study this fundamental question. The system faces a plethora of different sounds such as broad-band noise (rivers and wind), vocalizations (animal calls, speech), music, and random "packets" of noise (crackling fires and rain). Even when presented multiple stimuli at once, the system can separate, localize, and identify each source. Furthermore, it can infer parameters of each sound: roaring rapids in the distance vs a nearby creek, american vs british accents, and a clarinet vs a violin are easily differentiable. The complex process with which the auditory system analyzes these scenes is an ongoing hotbed of investigation.

Dedicated processing

The auditory system is a multi-tiered system in which sound information is processed in many stages before reaching the cortex. First, sound waves must be converted to neural signals. Pressure oscillations in the air are mechanically transmitted to the cochlea into fluid-solid waves which depolarize inner hair cells, in turn activating the auditory nerve (AN). The information conveyed through the AN then goes through several intermediate steps, detailed in Figure 1, before ending up in the cortex.

Throughout the system there is *tonotopy*: a mapping between physical space within a given brain region and the frequency that the region responds to; tonotopy is usually linear in the *log* of frequency (Pickles, 2008:various pages). Tonotopy originates in the cochlea, which contains a bank of bandpass filters with the base (where waves enter) sensitive to high frequencies and the apex responsive to low frequencies (Pickles, 2008: p37). Broadly speaking, this initial filtering step sets the stage for more complex frequency-time processing.

The ear is such a refined sensor that it generates a rich, yet challenging, data-stream. The cochlea has dynamic range of at about 6 sound pressure orders of magnitude (12 energy orders of magnitude), within which sounds are loud enough to be analyzed but don't immediately cause damage (Pickles, 2008: p276). Part of this dynamic range is provided by compression non-linearities in the cochlea and the rest by processing in the CNS (Pickles, 2008: p282). For lower frequencies there is also temporal information: the AN signals *phase-lock* to the waveform strongly up to 2 kHz, weakly up to 5kHz, and very weakly up to 12kHz (Pickles, 2008: p83-84). Phase-locking means that the action potentials (spikes) responding to a tone burst will preferentially occur at a certain point in the wave. No neuron can maintain a firing rate above a few hundred spikes per second, so an ensemble of AN neurons must be combined to guarantee that there will be spikes every cycle (Pickles, 2008: p82). Such high-fidelity temporal information complements the frequency selectivity of the cochlear filters, but in order to process this information the neural circuits later on must have very precise timing sensitivity.

Another factor that enhances capability yet complicates data analysis is the

external ear's “spectral notches” (narrow frequency ranges that are attenuated by interference created by the pinna). The location of the spectral notches depends on both sound source elevation and azimuth (Pickles, 2008: p290). Binaural (almost exclusively azimuth) timing and intensity further helps with localization (Pickles, 2008: p291). A spectral notch is easy to detect for a single sound, but when there are multiple sounds from different directions presented at once the notches get covered up. For example, sound X may be located in a direction associated with a notch at 3kHz, while sound Y is coming from a different direction associated with a notch at 4 kHz. If sound Y has energy at 3 kHz it can mask the X's notch, and visa-versa for X masking Y's notch. Dealing this this issue and combining both azimuth and elevation information, as in the case of separating sounds into a single percept, requires non-trivial processing.

The auditory system deals with this complexity with an intricate web of brainstem structures. The system has far more sub-cortical nuclei than andy of the other five senses (see Kandel, Schwartz et al, 2012 for an overview of the neural organization of the *six* senses). Figure 1 shows a simplified schematic of the pathways. Having so many structures below the (relatively slow) cortex allows an abundance of high-speed processing of temporal cues. The sub-cortical nuclei are arranged in a series-parallel progression, ferrying information through a variety of multi-step pathways to the cortex, with each step integrating contralateral (ear is opposite to the nucleus) and ipsilateral (ear is on same side as nucleus) information from the previous step(s) (Pickles, 2008: p182-3). The information gets broken up into multiple streams starting with the cochlear nucleus (CN), the first post-AN level in the pathway. In the CN, primary-like spherical bushy

units (mostly) preserve the AN raw data, octopus cells detect broadband onsets/envelope peaks, and fusiform cells detect spectral notches (Pickles, 2008: p157-162). The superior olive (SO), which takes input from the CN, measures and encodes both timing and level differences for localization (Pickles, 2008: p181). The next level, the lateral lemniscus, uses patterns of excitation and inhibition to differentiate left vs right sound locations (Pickles, 2008: p182). The Inferior Colliculus (IC) is the final major subcortical step in a sense: although the medial geniculate body, part of the thalamus, is between the IC and cortex, anatomical connectivity evidence indicates that "the cortex and medial geniculate body are grouped together as a functional unit" (Pickles, 2008: p194).

The auditory system also has a *descending* pathway (not shown in Figure 1). Information flows from the cortex all the way down to the cochlea, taking a similar route as the ascending pathway except backwards (Pickles, 2008: p239). The purpose(s) of this system are poorly understood. One mechanism may be gain control (Lyon, 1990). The cochlea's outer hair cells are tiny amplifiers that feed energy back into the wave as it is propagating from base to apex (Pickles, 2008: p125). The gain on this amplifier is very high, around 50 dB at low sound levels (Pickles, 2008: p143), but is reduced by stimulating descending neurons in the medial olivocochlear system (Pickles, 2008: p243). Spontaneous excitation of the cochlea can occur in certain types of tinnitus (Pickles, 2008: p132-134), which could represent a failure of a mechanism that normally allows high gains but prevents a feedback loop. Other functions of the descending pathway may be selective attention (Pickles, 2008: p251) and hearing loss protection (Pickles, 2008: p248), using gain-control to prevent amplification of unwanted or dangerously loud

stimuli.

In both the ascending and descending pathway, the Inferior Colliculus (IC) is a crucial, if poorly understood, hub that sits just below the thalamo-cortex (see Figure 1). Almost all the ascending auditory information passes through the IC, the exception being a small number of fibers that bypass several levels, jumping directly to the thalamus from the cochlear nucleus (Malmierca, 2002). The IC's central nucleus (ICC) receives the bulk of the ascending information from most of the lower structures, including the cochlear nucleus and the superior olive (Winer and Schreiner, 2005: p122 and Pickles, 2008: p187). Most of the descending pathway that originates from the the cortex targets either the medial geniculate body (MGB) or bypasses the MGB and ends up in the dorsal cortex and external nucleus of the IC (Winer and Schreiner, 2005: p241). This downward pathway is significant: in one study half of the IC cells were activated by cortical electrical stimulation (Winer and Schreiner, 2005: p234). The IC also talks to itself. There are a plethora of connections between the two IC's, as evident from the prominent wiring connecting the two “hemispheres” visible in a stained sample (Winer and Schreiner, 2005: p156). In addition, there are neurons, found mostly in the ICC and dorsal cortex, that project to neurons within the same hemisphere (Winer and Schreiner, 2005: p159-162, p166). ICC units, like units at many levels in the auditory system, usually have a well-defined frequency to which they are most sensitive (the “best frequency”, or b_f) and threshold (the minimum absolute dB-sound-pressure-level (dB SPL) that elicits a response) (Winer and Schreiner, 2005: p313). ICC units are also characterized by how they respond to stimuli at different sound levels (loudness) and frequencies near but not

at b_f; Figure 2 shows these main types of ICC units: Type I units only respond to tones placed very close to b_f, type V units have a wider tuning width, especially at higher sound levels, and the rarer type O units have more complex tuning properties such as responding to a narrow range of sound levels. Despite characterizing IC units' responses to tones and to several types of stimuli, there has been less success in painting a comprehensive picture as to what is the *function* of these units and the IC as a whole. One study showed that the IC units, acting as filters, provide more information to natural sounds than artificial ones (Attias, 1998). This is in line with the AN units being near-optimal encoders for a bank of natural sounds (Smith and Lewiki, 2006). Nevertheless, much research is needed in order to elucidate what this prominent structure is doing.

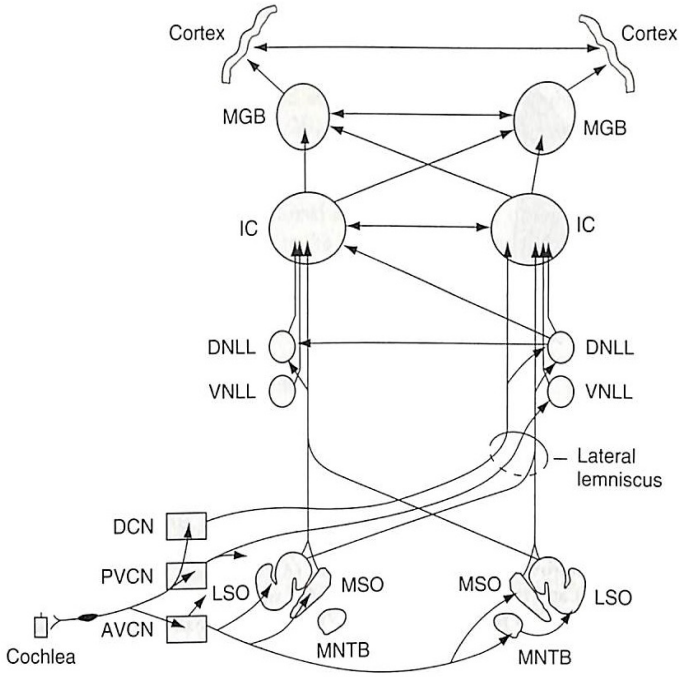

Figure 1: A diagram of the main ascending auditory pathways. Key: DCN, PVCN, AVCN = Dorsal, anteroventral, and posteroventral cochlear nucleus, respectively. LSO, MSO = lateral and medial superior olive, respectively. MNTB = medial nucleus of the trapezoid body. DNLL, VNLL = dorsal and ventral nucleus of the laterial lemniscus. IC = inferior colliculus. MGB = medial geniculate body (Reproduced from Pickles, 2008: p172).

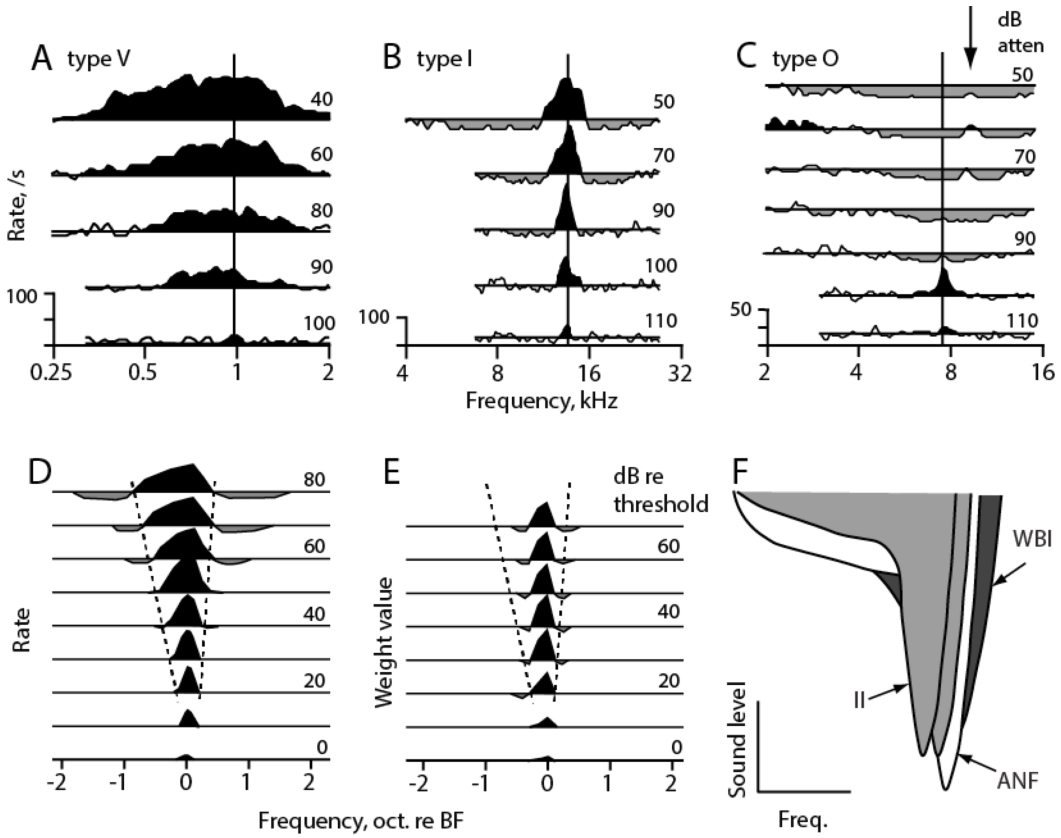

Figure 2: The response (spiking rates) to ICC neurons in decerabate cats.
A-E: each curve indicates the response (black is excitatory, grey is inhibitory) to tones at different frequencies for ICC units. In all examples, higher on the vertical axis is louder but "dB atten" in A-C is the *amount of attenuation* so louder is lower numbers.
F, inputs to DCN neurons, the excitatory area (white) is shadowed by inhibitory areas (gray). At the ICC level the pattern of excitation and inhibition will be even more complex.
(Reproduced from Young, 2010: p110).

Perception of pitch and harmonics

The auditory system assign a "pitch" to many sounds. A sound's "pitch" is defined here in accordance with American Standards Association (1951:p22) as the frequency of a tone that matches it on a scale of "low" to "high". For tones, by definition, pitch is identical to frequency, but there is no simple "rule" for calculating a pitch for complex sounds.

Sound waveforms that repeat exactly at a given period $1/f_0$ will usually have a pitch of f_0 (Cariani and Delgutte, 1996). These sounds are *harmonic* or *harmonic complexes* because the repetition constraint forces the Fourier components to be integer multiples of the *fundamental frequency*, f_0. Harmonic sounds are quite common in nature. Many animal calls (including human voice), musical instruments, and inanimate sounds are harmonic (see Fletcher, 2010 for an overview of the physics behind musical sounds). For clarity, we call the component with frequency kf_0 the *k'th-order component* or *k'th component* of a given *harmonic complex*. Sounds are only considered harmonic here if there are at least 2 components with reasonably small k-values. This restriction prevents tones and sounds with components at esoteric frequency ratios such as 300Hz:505Hz (f_0 would be 5Hz) from being considered harmonic. Orthogonal to pitch is the *timbre* of a sound. Rather than depending on f_0, timbre varies with the relative energies and phases of the components and differentiates a violin chord, piano key, or vowel of the same pitch (American Standards Association, 1951: p25). The auditory system uses both timbre and pitch information to represent harmonic sounds.

Harmonicity is important for sound separation and identification. The auditory system can pick out individual sounds from a “cocktail” consisting of multiple simultaneous sources (Bregman, 1990 and Darwin and Carlyon, 1995). A classic example of this is voice separation (Scheffers and Maria, 1983: p134). Each harmonic complex is perceived as a *single* auditory object with it's own pitch at f_0, even when the fundamental frequency is removed and cochlear non-linearities (which can generate energy at f_0) are masked (i.e. by loud low-frequency noise) (Licklider, 1954). This means that a wave

given by P = cos(2 pi 200*t*) + cos(2 pi 300*t*) will have a 100 Hz pitch even though there is no energy at 100Hz. Thus pitch for harmonic sounds is correlated to periodicity in the waveform's overall structure.

Pitch also exists for some non-harmonic sounds. Pitch for *mistuned complexes*, sounds that are identical to harmonic complexes except that all components have been shifted by some frequency d_f, depends on the degree of mistuning. Strictly speaking, the waveform of any sound repeats at a frequency given by the greatest common devisor (GCD) of the constituent frequencies. A precise GCD function is not useful, however, since it is infinitely sensitive to small changes. In practice, shifting a harmonic complex by d_f means that the wave's overall structure (the envelope) still repeats at f_0 but the wave itself is different at each envelope peak. The pitch, according to De Boers rule, begins to follow the individual components: for small d_f values the pitch is equivalent to a tone at $f_0 + wd_f$ with $w > 1$, but when d_f exceeds (approximately) $f_0/2$ the pitch will jump from $f_0 + wd_f$ to $f_0 - wd_f$ and will keep rising with a slope of k with further mistuning (Cariani and Delgutte, 1996). Finally, when d_f is f_0 the pitch ends up back at f_0 and the complex regains harmonically. These changes in pitch, despite maintaining a constant envelope frequency, demonstrate that the “fine structure” of the wave also contributes to pitch.

Auditory nerve responses to harmonic sounds

How does the auditory system transform a waveform into an ensemble of auditory objects, each with their own pitch and timbre? Pitch-detection starts in the auditory nerve. Electrophysiological studies have suggested that pitch is in part due to periodicity of AN

fiber spike trains (Cariani and Delgutte, 1996). With or without a fundamental, a harmonic waveform will repeat with a frequency of f_0. Envelope information is important to speech recognition (Heinz, 2009). A slowly-varying envelope will be explicitly encoded as a time-varying discharge rate in AN fiber spike trains. In addition, AN fibers use "fine-structure" information to encode the envelope implicitly as a cross-correlation between units with different b_fs *even if the envelope itself is removed* (Heinz, 2009). In both cases, extraction of the envelope is a non-linear process which adds a prominent component at f_0 to the Fourier transform of AN discharge rates, allowing pitch salience despite a missing fundamental.

Despite it's role in pitch detection, the auditory nerve does not show selectivity to harmonic sounds in terms of overall firing rate. Auditory nerve fibers act as band-pass filters for harmonic sounds, much as they do for tones (Cedolin and Delgutte, 2010). This filtering sets the stage for how processing at higher levels of the system will behave: harmonic components that can be differentiated from each-other (are *resolved*) can be processed in the frequency domain, but time-domain processing must be used for denser, unresolvable harmonic sounds. A simple measure of resolvability of a harmonic complex in the AN used in Cedolin and Delgutte (2010) is to look for peaks in the average discharge rate as a function at f_0 for an appropriate best frequency b_f. When $f_0 = b_f/k$ for integer k the unit will be driven strongly because a component of the sound is at b_f. For resolved complexes, when $f_0 = b_f/(k \pm 0.5)$, the response will be weak because there is no component at b_f and the components at $b_f \pm f_0/2$ are too far away from b_f to drive the unit (Cedolin and Delgutte, 2010). However, if the complex is "unresolved", the components

are so closely spaced (large b_f/f_0) that several of them land in the unit's receptive field and the response to $f_0 = b_f$ and $f_0 = b_f/(k\pm0.5)$, or anything in between, will be similar to the response to a tone at b_f (Cedolin and Delgutte, 2010). The response of a bank of AN units to a given sound as a function of b_f is a "blurred" version of the sound's frequencies; the degree of "blurring" limits the resolvability of individual components.

Resolvability can also be calculated *before* taking time-averages of firing rates. The "mean absolute spatial derivative" (MASD), defined in Cedolin and Delgutte (2010), is the difference in (expected) firing rates *at each instant in time* between adjacent AN fibers. To calculate the MASD at a given b_f, the absolute value of the differences between two fibers with b_f - ½ Δf and b_f + ½ Δf for a small Δf is evaluated at each point in time. These results are then averaged the $dt = 1/f$ ($1/f_0$) time interval for tones (harmonics) (Cedolin and Delgutte, 2010). The MASD is qualitatively similar to the average response. Sound energy at f will not only drive the units with $b_f = f$, there will also be a large phase *gradient* in the vicinity of f because the phase shifts by 180^o for a 2nd order system, such as the cochlea, when the resonant frequency crosses the driving frequency (Cedolin and Delgutte, 2010). The MASD is sensitive to on phase gradients, so it can be used as a surrogate for the average firing rate.

The resolvability of components based on the average firing rate vs that based on the MASD depends on the best frequency of the unit. At low b_fs neither metric indicates good resolvability, at intermediate b_fs the MASD metric had stronger resolution, but at higher b_fs the rate metric is stronger in model AN units (see Figure 3). At higher frequencies and lower sound levels more components were resolved overall by both

metrics (Cedolin and Delgutte, 2010).

Resolvability probably does not depend on the phases of the sound's components. Three types of phasing options: “COS” phase, where the phases of each component were the same, “ALT” phase, which was the same as the COS option but all the odd-order components are shifted 90^{o}, and “Schoreder” phase for variations in the envelope were minimized, yielded the same resolvability scores in model AN units (Cedolin and Delgutte, 2010). The phasing options determined the temporal structure of the sound-wave. Both the MASD-based and rate-based metrics showed very little difference in resolvability for the different phase options (Cedolin and Delgutte, 2010). Resolvability is primarily a function of the metric used and increasing function of b_f.

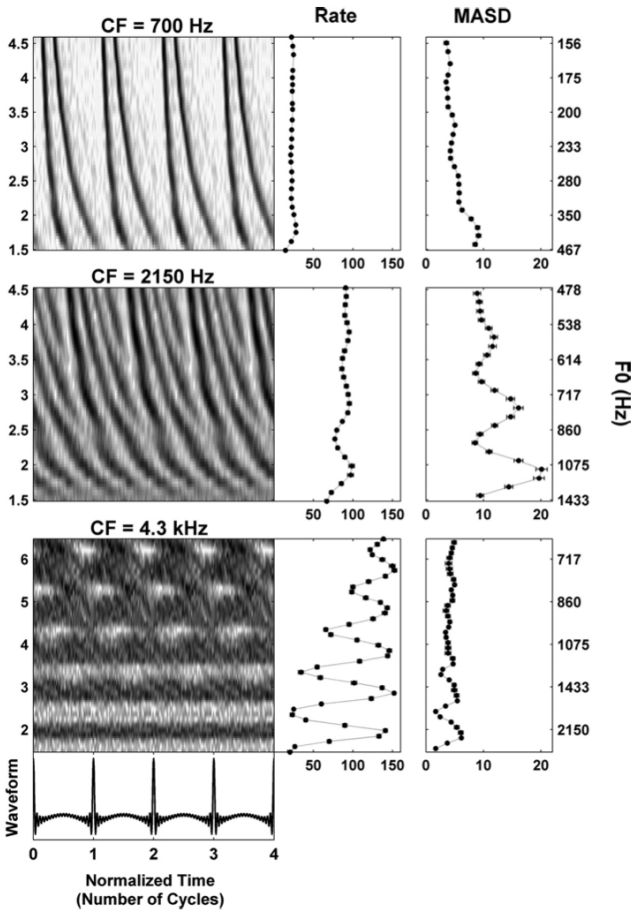

Figure 3: Response of three model AN fibers to harmonic complexes. The images on the left are the discharge pattern as a function of normalized time t/f_0 and the sound's f_0 (*higher* values of the harmonic number indicate *lower* f_0). The "CF" in the title is our b_f. The plots on the right are the rate and MASD of the fiber (the vertical axes is the same but is shown as a function of f_0). "Resolvability" means a pattern of alternating peaks and valleys on either the rate or MASD plot. The bottom plot on the left shows the sound waveform used for these plots.
(Reproduced from Cedolin and Delgutte, 2010: p12715).

Origen of pitch/harmonic-sensitive cortical units

The AN temporal-frequency representation contains all the information needed to separate and identify sounds, but most of this information is implicit and must be explicitly recoded later on. Harmonic/pitch sensitivity *is* explicitly encoded in the

primary auditory cortex (A1). In the awake Marmoset A1, “Harmonic template units” were found that responded much more to harmonic complexes than to tones *or* to mistuned complexes, which are harmonic complexes in which all frequencies were shifted by the same d_f (Feng, 2013, p48). Some examples were striking. Figure 4 shows a unit that had a much, much stronger response to harmonics than to tones, and Figure 5 shows very strong sensitivity to mistuning. Template units therefore encode the presence of a harmonic sound with a fundamental frequency (and pitch) of f_0.

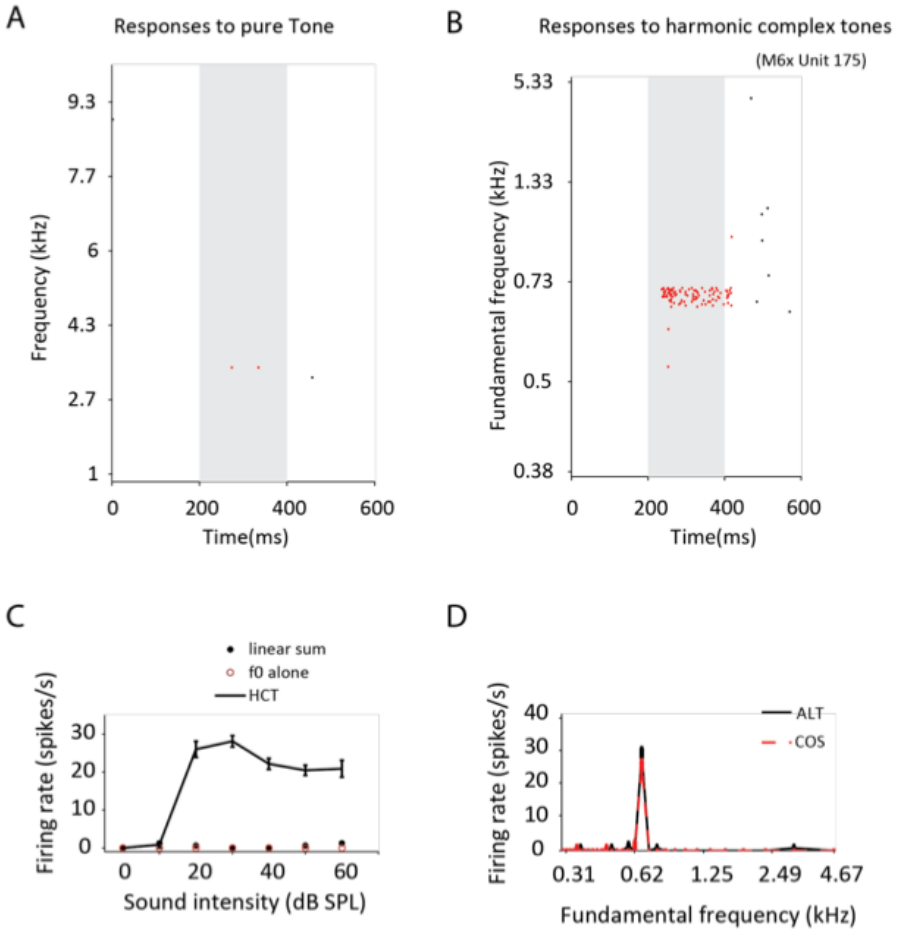

Figure 4: A unit with near-perfect selectivity for harmonic sounds.
A, The *dot plot* of responses to tones. Each dot is a single spike. Each row is a different presentation of a stimulus with a given f_0 (10 rows or so have the same f and then f "jumps"). The shaded region is the time period when the sound was presented. If we trust the three spikes to be indicative of a response, the b_f was 3.7 kHz.
B, Like A but with a harmonic complex with varying f_0. The complex's components cover the receptive field of the unit. Note the different frequency axes. The block of solid response corresponded to a single f_0 that strongly drove the unit, this was a b_f/f_0 of 6.
C, The response rate to harmonics (black curve) against dB SPL shows that, with the exception of very quiet sounds, the response did not depend strongly on sound level.
D, The response rate as a function of f_0. This unit did not show any phase preference between the COS and ALT phases. The tuning of the unit was sharper than this plot's resolution: only $f_0 = b_f/6$ yielded a strong response.
(Reproduced from Feng, 2013: p45).

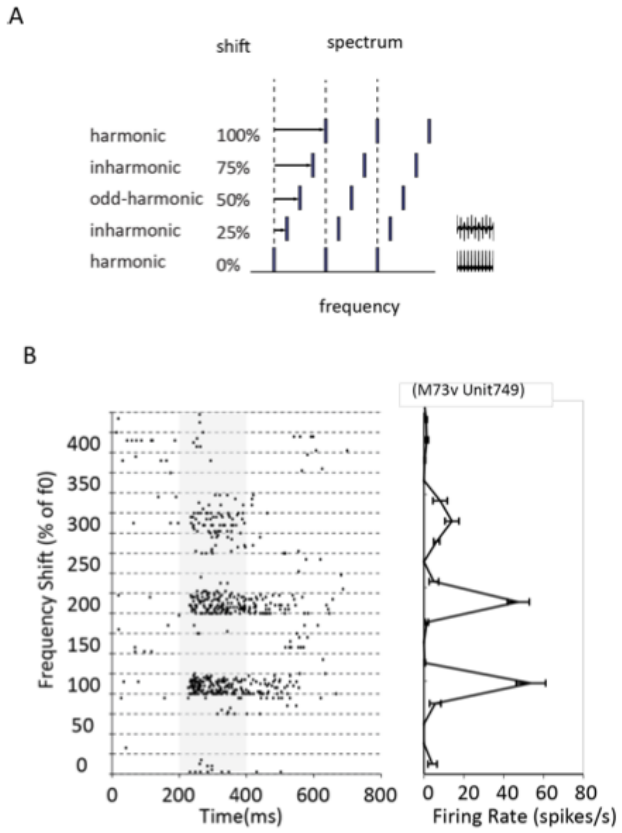

Figure 5: A different template unit that showcases the templates' much stronger response to harmonics than to mistuned sounds.
A, Schematic of mistuning a harmonic sound. Each bar represents a component. The tiny pattern at the right shows the sound waves.
B, The dot and rate plots much like Figure 4 B and D (the rate plot is turned sideways to align the frequency axes). The response was much weaker when the complex was mistuned.
(reproduced from Feng, 2013: p46).

How would such a selectivity come to be? In most of the stimuli used in Feng (2013), the sounds contained harmonics that were *resolved*, allowing for frequency-based processing. A spectral sieve behavior was found: template units took excitatory input from "ordinary" units tuned to $k_e b_{f0}$ and inhibitory inputs from units tuned at $(k_i \pm 0.5) b_{f0}$ for a subset of integer k_e and k_i values (see Figure 6). A harmonic sound with $f_0 = b_{f0}$ would land on all the excitatory inputs and energize the unit, while harmonic sounds with different f_0 *or* non-harmonic sounds would be mismatched and fail to energize the unit.

Pure tones would also poorly drive the unit because they would activate at most one input at a time.

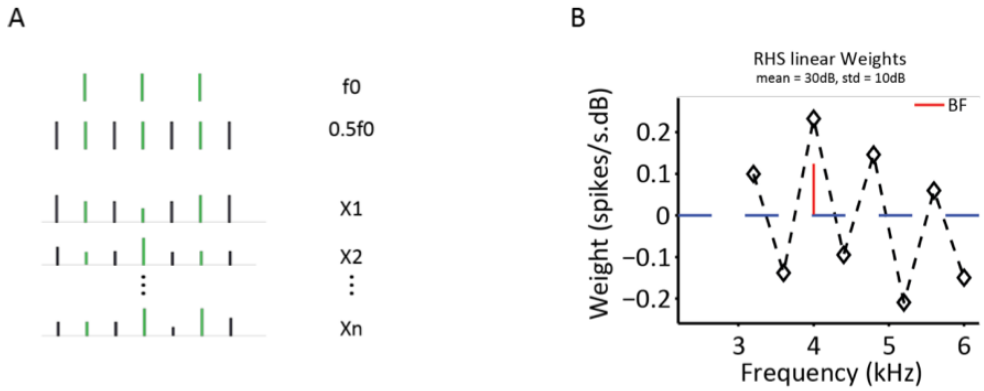

Figure 6: Random Harmonic Stimuli (RHS) were used to fit a linear weighting model to a harmonically sensitive unit (Feng, 2013: p36).
A, Example frequency-domain stimuli, longer bars denote higher energy. The fundamental frequency f_0, for which the random stimuli was based on, was the f_0 that yielded the peak response compared to any other harmonic complex (Feng, 2013: p11). In this case $f_0 \sim 800$ Hz.
B, The RHS weights. There were excitatory regions at 4, 5, 6, and $7f_0$ and inhibitory regions at 3.5, 4.5, 5.5, and $6.5f_0$; this alternating pattern is a "spectral sieve".
(Reproduced from Feng, 2013: p59).

A bank of such sieves could measure the timbre of *each* harmonic sound in a multi-sound stimulus. Each unit in the bank would have a given b_{f0} and be receptive to a narrow range of k values, effectively measuring a small portion of a single sound's components; inhibition could be used to help screen out distractor stimuli. The bank would consist of a grid of template units with different parameters that would create a 2D map of how much energy was at each k and f_0. The information needed to find the timber, how strong each component is, of a harmonic sound would be represented by a line of constant f_0 on this map. Such a model could be extended to include location information if the inputs to the grid units are direction sensitive, or even phase information if the inputs were sensitive to phases between components. In the marmoset auditory cortex, harmonic template units are sensitive to particular ranges of components as well as f_0

(Feng, 2013: p93). This suggests that timbre is identified in part by multiple units "sieving" a given harmonic complex.

The sieve method only would work for resolved harmonics. For dense, unresolved complexes, a time-domain method is necessary. Envelope modulation provides a simple method for dense sound *identification*. Consider a complex with components from $20f_0$ to $40f_0$ (k=20-40) and an f_0 of 100 Hz. This will excite AN fibers in the 1-2kHz region. The spacing is so close that the AN population sees the region as a continuous band and can't tell the spacing between components. If f_0 is halved and k doubled the range of excited AN fibers will be unchanged. However, envelope information can reveal f_0. If we have COS phasing, the envelop will have "peaks" that occur at a frequency of f_0 (Cedolin and Delgutte, 2010). For "Schoreder" phasing the complex becomes a series of very fast frequency sweeps; this will still create a (weaker) peak structure due the filtering in the cochlea (Cedolin and Delgutte, 2010). In both cases filtering smooths out these packets a little bit but this blurring does not prevent us from detecting them: when the frequency resolution of the AN fibers isn't high enough to differentiate individual components spaced f_0 apart the temporal resolution *will* be enough to resolve peaks that are $1/f_0$ apart. Although the envelope patterns in AN fibers reveal f_0 for COS and Schoreder phasing, ALT phasing introduces envelope peaks every $2f_0$ instead of every f_0 (Cedolin and Delgutte, 2010), which may "trick" the system into thinking that f_0 is doubled. A small set (4 out of 375 tested) of the units in the Marmoset auditory cortex showed sensitive to phasing to harmonic complexes (Feng, 2013: p70). These units, called "modulation sensitive neurons", seemed to prefer a specific envelope repetition rate (see Figure 7).

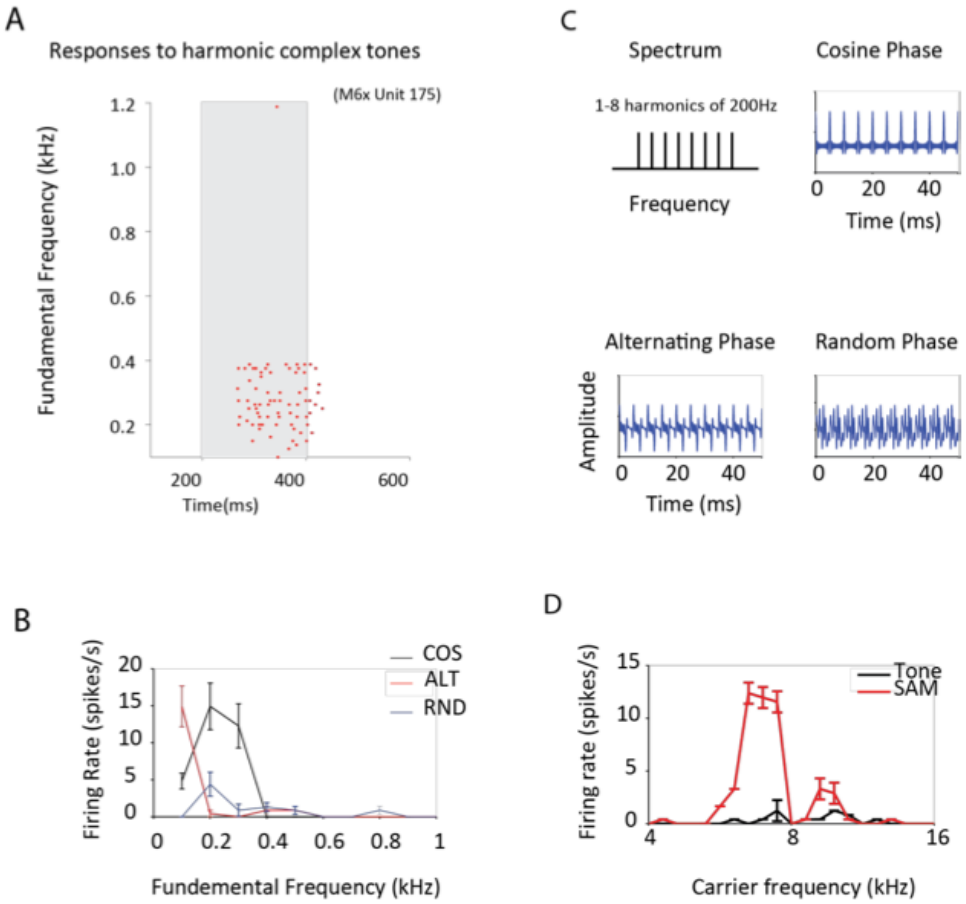

Figure 7: An example of a modulation sensitive neuron.
A, the dot plot as in Figure 4. The shaded region indicates when the sound was presented.
C, Schematic of the three phasing options. All have the same spectrum (top left) but the different options produce different time-domain waveforms.
B, firing rates to each phasing option. In all cases the unit seems to prefer a 200Hz envelope. The unit has a peak at about 200 Hz to the COS and random phasing and strong responses (possibly a peak) to a 100 Hz stimulus for the ALT phasing.
D, The response to a tone at a given "carrier frequency" that is modulated (has "beats") at 256 Hz. The response to the modulated signal is much stronger than that to an equivalent pure tone. Also, the carrier frequency is much higher than the preferred envelope rate.
(Reproduced from Feng, 2013: p89).

Neurons similar to the modulation-sensitive units in Feng (2013) have also been seen in the same preparation but with a click-train stimulus instead of harmonic complexes. In (Bendor and Wang, 2010), a "modulation sensitive unit" (MSU) was essentially a unit that responded to narrow-band click trains with a given repetition rate

but responded to tones at a different (usually much higher) frequency (see Figure 8).

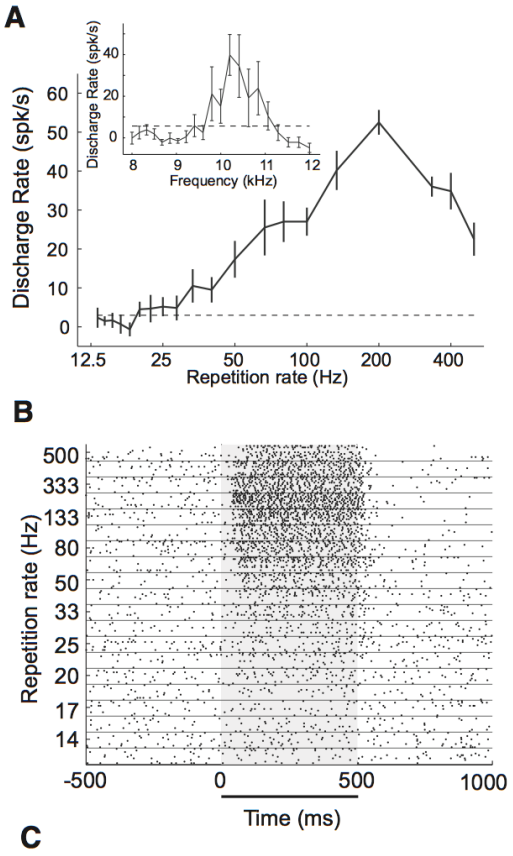

Figure 8: An example of a modulation sensitive unit (MSU).
A, The firing rate of an MSU that responds to best pure tones around 10.5 kHz (inset) but responds best to click trains with a repetition rate of 200Hz. The dashed line is the spontaneous discharge rate.
B, The dot plot (as in Figure 7, except this time with horizontal lines delineating blocks of identical stimuli) of this unit, showing greater sustained responses near a 200 Hz repetition rate. This neuron does not phase-lock to the pulse-train but did show a rate preference.
(Reproduced from Bendor and Wang, 2010: p1812).

Another population of units in Bendor and Wang (2010), called "pitch sensitive" units, showed sensitive to temporal *regularity* in click trains. Unlike the the MSUs, which responded at the same rate whether or not the timing of the clicks was jittered, the pitch sensitive units *dropped* in firing rate when there was jitter (Bendor and Wang, 2010). Regularity is harmonicitiy: jittering the click trains destroyed the harmonic structure

because it disrupted the temporal repetition (see Figure 9). In the example in Figure 9, the click trains were a harmonic complex that in the frequency domain that peaked at k=30 and has significant energy in the range of k =25-35; this is well beyond the AN resolvability limit in Cedolin and Delgutte (2010). Any pitch sensitivity must arise from either extreme sharpening in the frequency domain (which is unlikely), or temporal processing. Furthermore, any temporal analysis must be subcortical since the click rate of 100 Hz is above the 50Hz fusion threshold of cortical neurons (Bendor and Wang, 2010).

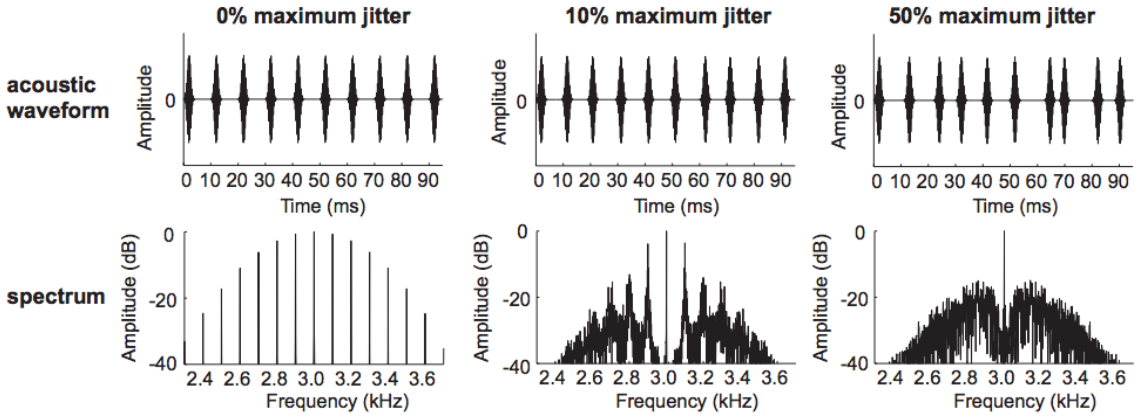

Figure 9: Jittered and unjittered click trains. The top is the waveform and the bottom is the normalized Fourier transform.
(Reproduced from Bendor and Wang, 2010: p1819).

Detecting temporal regularity is useful for *separation* of overlapping dense harmonic sounds. As in the case for resolved harmonics, the information is implicit in AN rates and must be extracted. Suppose there are two sounds with COS phase with $f_{0,a}$ = 200 and $f_{0,b}$ = 256 Hz that cover (almost) the same range of frequencies and are both composed of unresolved harmonics. The AN fiber's response will be a superposition of a pulse train at 200 and 256 Hz. The “modulation”, or inter-pulse interval of the combined train have any value from 0 to 5 ms but there will be a more consistent periodicity at 200

Hz and 256 Hz (the Fourier transform of the pulse train will have prominent components at these frequencies). In dense stimuli the AN is giving temporal information (the arrival times of the packets), so any Fourier-like analysis has to be done in software rather than hardware. At some level in the auditory pathway there must be circuits that pick out the real periodicity at at 200 and 256 Hz and ignore the other (spurious) inter-click intervals.

Limited progress has been made as to *how* sensitivity to temporal regularity arises. The "cancellation model" in Cheveigne (1998) proposes a series of delay lines that act as coincidence detectors: they only fire if they get two spikes separated by the correct time interval Δt, serving as a detector for periodicity. Inhibitory coincidence (the circuit fires whenever it get a spike *unless* there was a spike Δt beforehand) were also suggested and may be useful to prevent cross-talk when multiple harmonic sounds are present, but the details of how a periodicity detector would be built are poorly understood (Cheveigne, 1998).

Another effect that can occur with closely-spaced components for different sounds is "beats". If two components from different sounds are at similar dB SPLs and are both in receptive field of a given fiber (i.e. $k_a f_a \approx k_b f_b \approx b_f$), the interference will cause the fiber's response to have a beat frequency with $f_B = k_b f_b - k_a f_a$. The 50 Hz cortical fusion threshold (Bendor and Wang, 2010) can easily be higher than the beat frequency, obviating the need for the higher speed sub-cortical processing to "figure out" that multiple sounds occupy the given frequency band. Since higher beat frequencies are also possible, beat information would complement, but not replace, subcortical processing.

The spectral sieving of harmonic template units in Feng (2013) and the temporal

periodicity sensitivity of the units in Bendor and Wang (2010) are probably two different populations of units. The pitch units tended to have lower b_fs than the MSU's and were concentrated in a "pitch center" in the cortex (see Figure 10). On the other hand, the template units were found with a similar, slightly higher, spatial and frequency distribution as the non-harmonic units (see Figure 11). The predominance of temporal sensitivity at low frequencies, which was more applicable for the click-train stimuli, is not surprising since lower frequency allows AN phase-locking (Pickles, 2008: p83-84) and has (relatively) lower spectral resolvability and higher temporal resolvability (Cedolin and Delgutte, 2010). The pitch units filled this low-frequency niche while the harmonic units operated across the bulk of the hearing range.

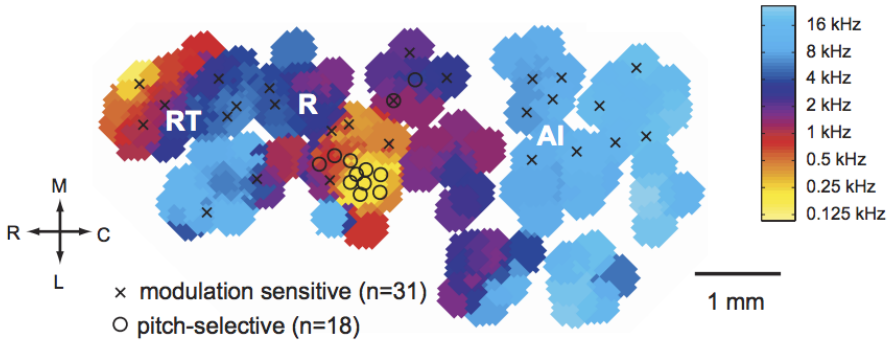

Figure 10: Locations of the modulation and pitch-selective units in a marmoset cortex. Left-Right is rostral-caudal, up-down is medial-lateral. The modulation sensitive units are spread throughout the cortex but the pitch-center is concentrated in one of the regions of low-b_f units. Three sub-regions of A1 are labelled: AI: Auditory area I, R: Rostral area, and RT: Rostro-temporal area. The pitch center was near the boundary of the core and belt areas of the cortex.
(Reproduced from Bendor and Wang, 2010: p1816).

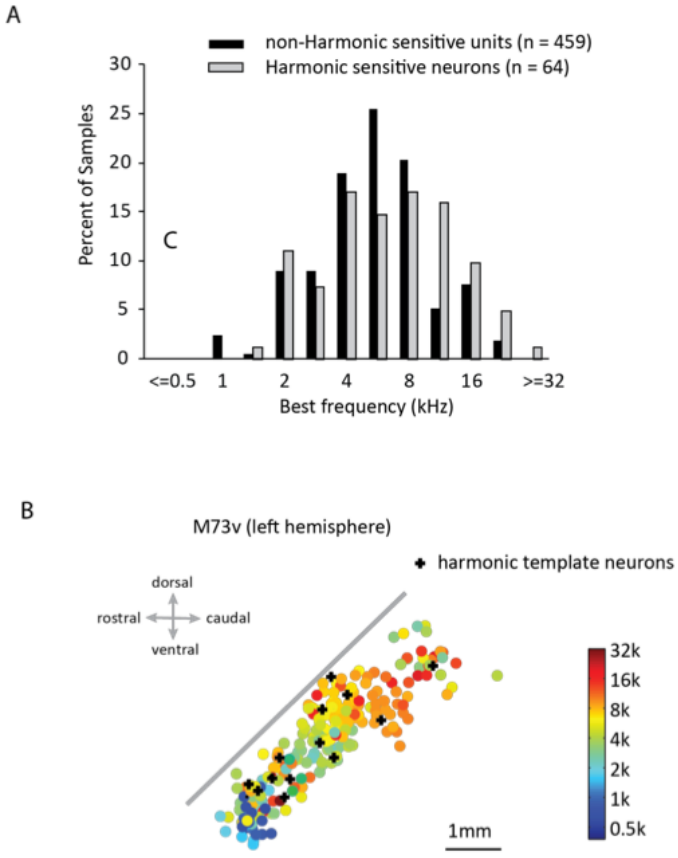

Figure 11: The response and location of harmonic template neurons is similar to all other units.
A, The range of frequencies of harmonic template units and non-template units.
B, the physical location of the template units in a single Marmoset's auditory cortex (A1), superimposed on a map of the b_fs for all units. The other two animals had different distributions of cortical b_fs but the template units still were found across the entire region. *The color-bar is different from Figure 10*.
(Reproduced from Feng, 2013: p52).

Motivation

The goal of this study is to elucidate the process by which a pitch/harmonically selective cortical unit arises. Is this sensitivity a de-novo cortical phenomenon or are there units in the Inferior Colliculus (IC) with the same properties? If there are no such units, are IC units useful in *constructing* a harmonically sensitive unit? This applies for both the time-domain selectivity described in Bendor and Wang (2010) and the frequency-domain selectivity in Feng (2013). Besides overall firing rate, are there any other metrics that show harmonic selectivity (to different f_0s and/or mistunings) beyond what is expected based on how IC units are known to respond?

Methods

Experimental preparation

The methods were similar to (Feng, 2013: p9). Sound presentations and single-unit recording were performed in a double-walled soundproof chamber (IAC Acoustics). Single units were isolated and recorded from one marmoset (young adult male, about 2.5 years old when training begun). Recordings were done from both inferior colliculi (ICs), however the data used in the analysis came from the left IC; preliminary datasets not useful for the modeling were taken from both the left and right IC.

Details of the chronic preparation can be found in Nelson (2009) or Lu (2001). Briefly, the animal was trained to sit still in a primate "chair" by gradually increasing the length of restraint in the chair up to about 3 hours. Once trained, it was implanted with a dental cement "head cap" fixed to the skull under general anesthesia and allowed to recover. A 1mm crainotomy was drilled through the dental cement and bone to access the brain; old holes were sealed and new holes were drilled when excessive tissue growth became problematic. A sterile bandage and Examix "pink rubber" filler protected the interface between the cap and the skin and the access hole to the brain, respectively.

Sessions lasted just under 4 hours on average (typically they ranged between 3 hours to 4.5 hours chair-time). The awake animal was chaired and head-fixed for the duration of the procedure. Although this was an "awake" preparation, there was occasional sleeping but no change in neural response (apart from a reduction of motion-

induced firing modulation) was noticed.

A Tungsten micro-electrode (A-M Systems, 5MΩ) was inserted into the IC with a hydraulic micro-drive (Kopf). The electrode took a path from superior to inferior and from lateral to medial. This path went through non-auditory parts of the cortex and thalamus before entering the IC. When the electrode was in or near the IC, a background response of several units (a *hash*) was heard. The properties of the hash allowed us to determine if we were in the ICX or the ICC: In the ICX the hash was usually stronger to noise and in the ICC it was stronger to tones. Also, in the ICC the hash featured a tonotopic progression of tuning from low to high frequencies don the track. In both the ICX and ICC, hash and neurons were found with various search stimuli, typically tones and broadband noise with sound levels in the range of 40-50 dB SPL. Harmonic complexes were occasionally used at first but were not found to elicit any more response than tones or noise.

Electrode waveforms were band-pass filtered, typically between 400-5000Hz, but sometimes this was adjusted as needed for good isolation. Isolation was evaluated by listening to the filtered waveform to make sure there was only one type of "click" present, by ensuring that inter-spike intervals were at least 0.8 ms apart (assuming a 1 ms absolute refractory period and accounting for jitter in the triggering), and by ensuring spikes could be cleanly differentiated from background noise. Spikes were detected with a Schmitt trigger (usually with a timeout of <0.2 ms) with a small hysteresis. Only the spike times were recorded.

Sound generation

Sounds were generated on the computer (MATLAB), converted to analogue (RP2, PA5, Tucker-Davis Technologies), amplified through a Parasound amplifier, and played with a Kef model SP3724 speaker. The sample rate was initially 100kHz but then changed to 200kHz to further minimize sampling artifacts. The speaker was front of the animal and 1 m away.

Stimuli design

The data consisted of multiple stimulus *maps* taken for each unit. Each map was a sequence of related sound presentations (tokens). Tokens in a given map were presented back-to-back while the spike times were continuously recorded, but there was time-gap that was typically 5-20 seconds in-between maps for which nothing was collected and no sound was presented. Each token contained a sound 300 ms long (starting and ending with a 10 ms cosine ramp) followed by 700ms of silence (this was later changed to 150 ms and 350 ms, respectively). The series of tokens explored the effects of varying a single parameter (such as f_0 in a harmonic complex). In all cases, energy above 40kHz, taken to be the hearing range limit of the animal, was cut off to avoid Nyquist aliasing. The following types of maps were taken:

- ***pure*:** (see Figure 12A). A sequence of tone bursts with frequency, f, varying logarithmically. The range of the map was determined by the width of the receptive field, and typically was about 1-2 octaves.

- ***hm*:** (see Figure 12B): A sequence of harmonic complexes with varying fundamental frequency f_0. Each token had components at kf_0, $k = \{1,2,3,...\}$. Usually several *hm* maps were taken, in total covering a range of f_0 from about $b_f/10$ to about $1.5b_f$.
- ***tsh*:** (see Figure 12C): A mistuned harmonic complex. Each token contained components with frequencies at $d_f + k f_0$, where d_f was the mistuning and a range of integer $k \geq 0$. The value of f_0 was fixed and d_f varied. Both the *hm* the *tsh* components were linearly-spaced in the frequency domain for any given token. Figure 12 shows schematics of the *pure*, *hm* and *tsh* maps.
- ***RLV*:** a map that presented a tone at a fixed frequency *f* while varying the sound level.
- **stretching tone complex**: (not used in the data analysis): A sequence of sounds each with linearly-spaced components, generated by fixing the the center frequency while varying the spacing f_0. The center frequency was, at all times, the average of the frequencies of each component and was almost always placed at the units b_f.
- **wanderer**: (used for exploration only, not used in the data analysis): A harmonic complex with fixed f_0 and with a wanderer added: an extra tone that moved through the complex and was usually at a louder sound level.

Exploratory study and final stimuli design

We initially applied a wide variety of stimuli to units before narrowing the

stimulus protocol. We varied parameters such as the stimuli used, number of tokens, and number of components over a wide range in order to find the most efficient protocol and to search for any unusual/unexpected effects on the unit. The data from the exploratory phase proved too heterogeneous to study but it was useful for designing a specific three-step process: Obtain the best frequency (b_f) and threshold, find the response to *pure* and *hm* at a fixed dB above threshold, and obtain *tsh* maps at the same per-component dB level with $f_0 = b_f/k$ and $b_f/(k+0.5)$ for integer k.

For the data-collection protocol, the *hm* and *tsh* were initially set to 7 components, which was sufficient in most cases to cover the entire unit's receptive field. However, this was later changed so that *all* cases are covered with a safety margin, in order to prevent edge-effects. The relative phases of the *hm* and *tsh* components were set to minimize variation of the envelope. For the *hm* this is given by:

$$\phi_k = \frac{\pi}{n} k^2 \tag{1}$$

Where Φ_k is the phase angle of the k'th component (i.e. the $\cos(2\pi t - \Phi_k)$ term in the waveform) and n is the number of components in the *tsh* or *hm* map. This formula was based on the phasing in Schroeder (1970) and has almost identical properties.

For the *tsh* we generalized this formula slightly:

$$\phi_k = \frac{\pi}{n}\left(\frac{f_k}{f_0}\right)^2 \tag{2}$$

Where f_k is the *frequency* of the k'th component.

Like any phasing option, this didn't abolish temporal structure: instead, the sound became a very fast frequency sweep, see Figure 12. We determined that sweeps were less

of a problem than sharp peaks in the envelope since strong AN envelope sensitivity was observed in studies such as Bendor and Wang (2010).

Any stimuli taken after the exploratory study (except the *RLV*s) were *interlaced*: instead of sequentially going from "low" to "high" stimuli (i.e. F_1,F_2,F_3 … F_{100}), we swept across the range of *F*-values 10 times, each "sweep" covering slightly different frequencies: (F_1, F_{11}, F_{21},...F_{91}, F_2, F_{12},...F_{92}, F_3, F_{13}, F_{23} …, …, F_{90}, F_{100}); *F* being *f* in the *pure* and f_0 in the *hm* and *tsh* maps. We then sorted the tokens from low *F* to high *F* for the data analysis. Interlacing reduced the effects of adaptation. Also, any several-token artifact, such as excitation from animal-generated sounds, would be broken up instead of creating a coherent "lump" on the map.

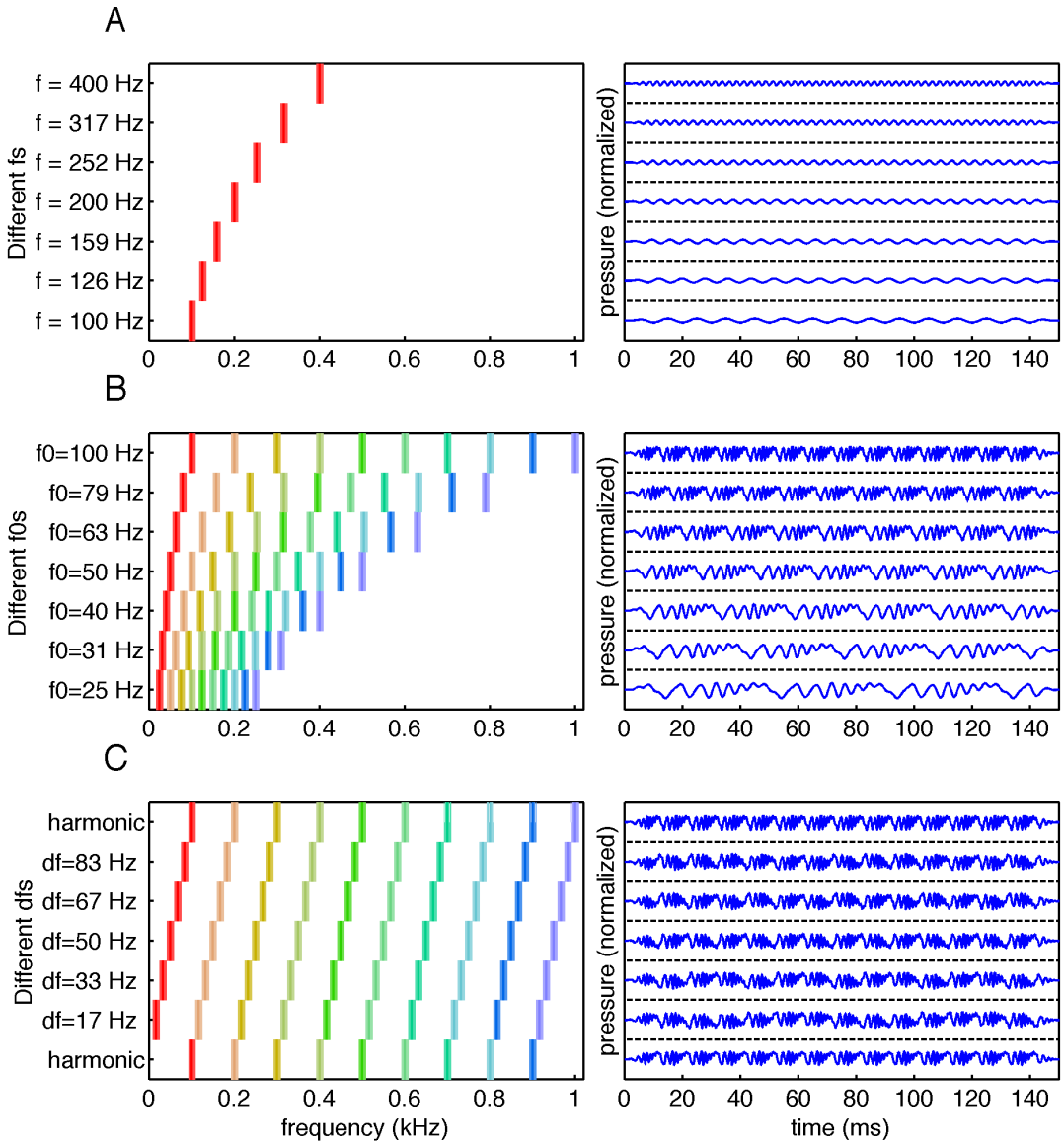

Figure 12: *pure*, *hm* and *tsh* map stimuli. On the left are the sounds, in the idealized frequency domain, for 7 tokens in a map. The lowest frequency component is red and the highest one is violet. The frequency scale is linear. The right plots show the waveform, for each token, with the same duration and ramping as our stimuli (150ms and 10ms, respectively). These tokens are around the lower limit of the frequency range actually used; this made the waves visible.

A, a *pure* map with f varying logarithmically from 0.1-0.4 kHz.

B, a *hm* map with 10 components, with f_0 varying logarithmically from 0.025-0.1kHz.

C, a *tsh* map with f_0 = 0.1kHz and d_f, the frequency of the lowest tone, varying linearly from 0 to 0.1kHz. There are 10 components (except when d_f = 0 when there are only 9 non-zero-frequency components). When d_f = 0 or d_f = f_0 , the stimulus is harmonic with fundamental frequency f_0. Note that the phasing works well in keeping the envelope smooth for both the *hm* and *tsh* maps.

Best frequency, threshold, and map collection

The first step after isolating a unit was to determine it's basic tuning properties because the unit's tuning determined the parameters used in the *hm* and *tsh* stimuli later on. As units were being stabilized we measured approximate tuning properties by manually testing tones at different frequencies.

Next we built a *response map* out of multiple *pure* maps at different dB SPLs, usually spaced 10-20 dB apart (5-10 dB apart near the approximate threshold). Each *pure* map consisted of 100 logarithmically-spaced *f* values designed to cover the excitatory region of the neuron as well as any nearby surround-inhibition. For the later units an *RLV* (with a soft to loud token progression but no interlacing) was taken for tones at b_f to obtain a more precise threshold; see *Firing rate and "spont"* and *Measures of tuning width and selectivity* for how a response map was visualized and analyzed. The response map was used to manually estimate the b_fs and thresholds. For units with no RLV available, the threshold was the lowest dB SPL *pure* map that a had a peak (a "peak" being a clearly elevated response to several nearby tokens that was unlikely to be a statistical anomaly). When there was an *RLV* map available, the threshold was defined as the lowest dB SPL for which there was a increase over several tokens above the silent level.

Once the b_f and threshold were determined, we measured the response to *pure* and *hm* maps at a sound level for which *each component* was at a constant level 30dB above the unit's threshold (this was later changed to 40dB above threshold). Like the *pure*, the *hm* used 100 tokens and had a logarithmically-spaced f_0. The b_f and threshold analysis

done during the experiment was redone more carefully offline so the exact value changed slightly.

The final step of data collection, taking *tsh* maps with $f_0 = b_f/k$ and $b_f/(k+0.5)$ for integer k, required a precise value of b_f. However, b_f is a function of both the sound level and the "crowdedness" of the stimuli. In anesthetized cats, Yu and Young (2013) used tones and "random spectral shapes" (RSS) to measure the response at a variety of levels. RSS, unlike tones, measures the incremental changes of the neuron's response *in a noisy environment* as perturbations are made to the sound (Yu and Young, 2013). The tone response typically broadened and shifted to lower frequencies at higher sound levels but the RSS response tended to be narrower and relatively independent of dB (Yu and Young, 2013). Because we were concerned with the response as harmonics are mistuned, and harmonic sounds are multi-component, we were more interested in the RSS-like behavior than the tone behavior.

We used *hm* stimuli to estimate the best frequency b_f. The response rate to the *hm* usually had peaks at $f_0 \approx b_f$ (the 1^{st} order peak), $f_0 \approx b_f/2$ (2^{nd} order), $f_0 \approx b_f/3$ (3^{rd} order), and so on. The first few-order peaks, typically to $\approx 3^{rd}$ order, were strong and very well isolated, with "valleys" in between the peaks at near-silent response levels. Higher order peaks, on the other hand, had less contrast and/or were weaker. We then selected a peak corresponding to an intermediate *reference order*, r, where the peaks were were clearly defined but not completely isolated from each-other. This intermediate choice of r meant that the components were interacting (multiple components in the receptive field of the unit) but that the stimulus wasn't so dense that the unit couldn't distinguish changes in f_0;

r was usually 3-5 but was lower when ≥3rd order peaks weren't strong and well-isolated. We then approximated b_f by calculating the "harmonic b_f", given by: $b_h = rb_{f0,r}$, where $b_{f0,r}$ is the f_0 of the *r*'th order peak. The difference between b_f and b_h was usually small, in about half of the units they were within 4% percent of each-other (see *Tuning properties*). However, a 4% difference was still enough to affect the location of the higher order peaks. Although b_h may vary with sound level because the stimuli were less "crowded" than the RSS stimuli, we didn't test for this because we collected very little *hm* data at multiple sound levels for a given unit.

After b_h selection, we presented a series of *tsh* maps at the same per-component dB level as the *hm*. Each *tsh*, like the *pure* and *hm* maps, had 100 tokens testing different values of d_f. However, the d_fs were on a *linear* scale and varied from 0 to f_0. We only needed to take up to $d_f = f_0$ because the *tsh* is a "circular" map: if $d_f = 0$ or $d_f = f_0$ (bottom or top row of Figure 12C), the position of each component will be the same except for the two "end-points". The *tsh* data was typically taken at integer and half-integer orders (meaning that the f_0 in each map was set to integer and half-integer fractions of b_h); occasionally a quarter fraction was tested. If time permitted, every integer and half-integer order from 1 up to the point at which the unit lost selectivity (less than about a 25% variation between the peak and valley of the *tsh* map) was tested.

Interleaved with acquisition of the *tsh* maps, which were usually the bulk of the data, occasional *pure* maps were sometimes taken (at the same per-component sound level of the *hm* and *tsh*). Having multiple acquisitions of the exact same map, spread out in time, allowed us to verify that the triggering stayed stable and that the unit wasn't

changing it's response properties. Also, the acquisition of multiple replicates reduced noise in the *pure* data.

Data Analysis

Data were analyzed both on a per unit and on an overall summary basis. MATLAB code was used for analysis and visualization.

Firing rate and “spont”

The firing rate for each *map* was calculated automatically. The *rate* for a particular token was the number of spikes in the time interval {10 ms, sound duration + 10 ms}, divided by the length of that interval. This gave us an estimate of the average response to the sound; the 10ms lag was an estimate of the combined latency of the speaker's mechanical impedance (which should be rather small), sound propagation through the air (about 3 ms across a 1 m distance) and signal transmission in the auditory pathway leading up to the IC (about 6 ms, Winer and Schreiner, 2005: p2). The *spont* was the number of spikes after ½ way through the token, divided by the length of that interval; this estimated the spiking rate in the absence of sound. These metrics weren't perfect: the neuron tended to have a stronger response right near the start of each token, and sometimes the spont was affected by the sounds as well. However, this still was a simple, objective measure that captured much of the behavior. Figure 13 shows the rate and spont calculation.

The rate and spont were smoothed with a 5 token median filter before extracting any statistic that was sensitive to individual outliers (i.e. the peak response). Calculating the smoothed rate for the two tokens nearest to the edge of the map would require sampling points off the edge. Instead, for these boundary cases we used the median of 5 points nearest the edge. Doing this instead of truncating the window meant that we always sampled 5 points regardless of the boundary conditions. This also meant that at least 3 points near each edge took on the same smoothed value. Figure 13 illustrates the result of smoothing. Acquiring the data with interlacing was vital for the robustness of any smoothing method since it made it much less likely that artifacts would clump together in a given filtering window.

In most cases we were interested in the *induced rate*, which is the spont subtracted from the rate. When there is no stimuli the rate should equal spont and the induced rate should be (on average) zero. We always applied smoothing *before* subtracting spont to calculate the *smoothed induced rate*.

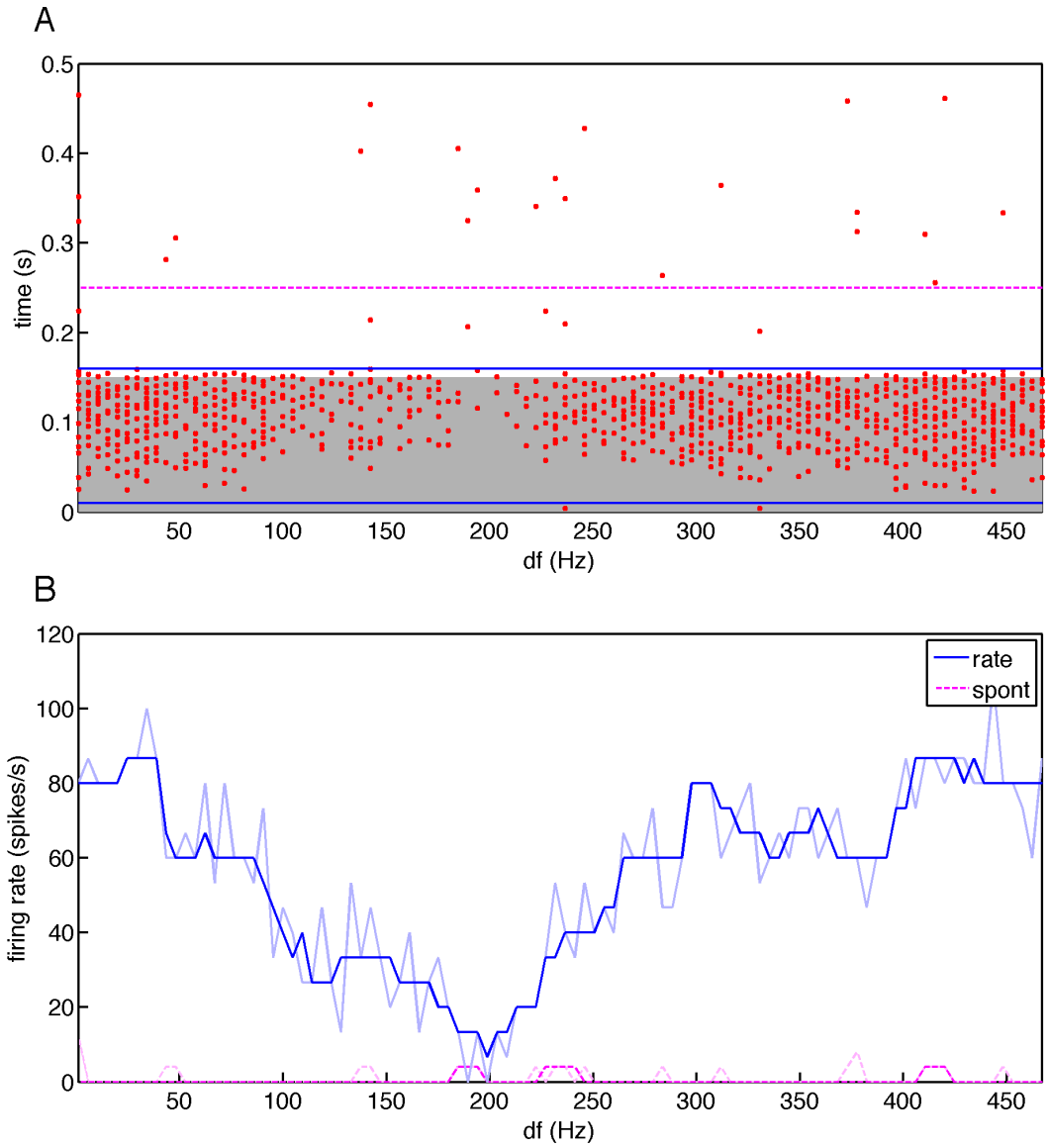

Figure 13: Calculating and smoothing the firing rate for a *tsh* map for a unit with b_f = 1.37 kHz, b_h = 1.4 kHz, and a 15dB SPL threshold. The firing rates are plotted against d_f. This *tsh* had f_0 = 467Hz, which was 1/3 of b_h (3rd order). There was a component of the stimulus at b_h when d_f = 0 or d_f = 467Hz, as expected the response was stronger in the vicinity of these d_f values.

A, The "dot plot" as in Figure 4, except with time on the vertical axes. We don't have groups of identical tokens as did Feng (2013), instead we used median filters to estimate properties such as peak firing rate. The shaded region is the time period for which the sound was presented. The two blue solid horizontal lines (which are 10 ms above the bottom and top of the shaded region) are the time window between which spikes are counted toward the driven rate. The spikes *above* the magenta dashed line at (more than 0.25 s into each token) are counted toward the spont.

B, The un-smoothed (faded lines) and smoothed (stronger lines) rate and spont calculated from A. The spont was very low in this unit and possibly was further inhibited on tokens that drove the unit strongly.

Measures of tuning width and selectivity

We used two measures of tuning. To measure the tuning width of the *pure* maps, which almost always consisted of a single, well-defined peak, we use the "full width at half maximum" (FWHM). The tuning width was the frequency difference between the edges of the region for which the smoothed induced rate was larger than ½ of it's maximum; Feng (2013: p12) used a 20% cutoff, but a 50% was more robust to noise.

The tuning quality, for each unit, was given by:

$$Q_{40} = \frac{b_f}{\Delta f} \tag{3}$$

Where Δf is the bandwidth for a *pure* map 40 dB above threshold, in accordance with Ramachandran et al (1999), except that we used the full width at half-maximum (FWHM) instead of the width of the entire excitatory area since the FWHM was more numerically stable to automatic analysis. A higher Q_{40} unit is more narrowly tuned and more sensitive to frequency differences near b_f. Since we usually didn't have a *pure* map at 40 dB above threshold in most cases, we interpolated between the nearest two *pure*s, on a log scale, as illustrated in Figure 14. For high-threshold units we didn't have a *pure* above 40 dB above the threshold (as that would be too loud to comfortably present to the animal), in those cases we selected the loudest *pure* and did not interpolate.

The *pure* map's rate and spont at each sound level can be visualized as a series of curves placed at different heights, which is called a *response map*. Figure 14 shows the response map along with the tuning width and Q_{40} calculations.

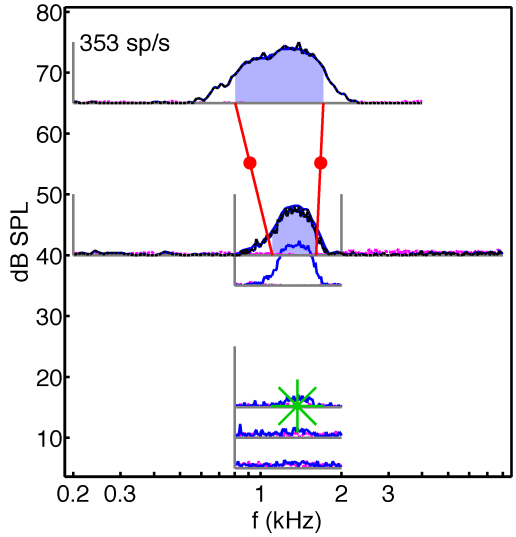

Figure 14: The selection of b_f and threshold and the Q_{40} calculation for the unit in Figure 13. The Q_{40} was 1.84. No RLV was taken for this unit. The response map is shown using a collection of subplots in a similar manner as Figure 2. Each subplot, indicated by the thin grey axes lines inside the main plot, is the firing rate (solid line) and spont (dashed line near bottom of each subplot) to *pure* maps as a function of frequency, as calculated as per Figure 13. The dB-position *of the floor* of each subplot is the dB SPL of the given map. The height of each curve *above the floor* is the firing rate. Every subplot is on the same scale: the global maximum of 353 spikes/s (see label near top left) is set to 10 dB on the graph. The star at 15 dB is the unit's threshold. The shaded regions at 40 and 65 dB indicate the regions within which the smoothed induced rate was at least ½ of its maximum for the corresponding sound level. To calculate Q_{40}, we needed the tuning range at 55dB. The minimum (maximum) frequencies at 55dB were estimated by interpolating, linearly in the log of frequency, between the frequency of low frequency edge (high frequency edge) at 40 and 65 dB. The lines connecting the edges of the shaded regions show this interpolation process and the dots are the interpolated frequencies themselves that were then used in eq. 3.

Another measure of tuning selectivity, more appropriate to "circular" maps, is the *vector strength*, which has been and still is used extensively in order to quantify periodicity in biological responses (Hemmen, 2013). Vector strength is applicable whenever the independent variable is "circular" (has a "wrap-around" effect in which two different values correspond to the same parameters). The *tsh* maps were circular because that tokens with $d_f = 0$ were identical, except for the boundaries of the stimulus, to tokens with $d_f = f_0$. Since the boundaries were usually outside of the receptive field, the response was almost the same at either token. The *hm* was also circular in the sense that the tokens with $f_0 = b_h/(k \pm 0.5)$ for a given integer k puts the unit's b_h in-between components, while

the token with $f_0 = b_f/k$ will put a components at b_h. This is not a perfect circle since the spacing between components was different at $b_h/(k - 0.5)$ and $b_h/(k + 0.5)$. However, the difference in spacing between these two endpoints is small at moderate and high orders.

We placed our response on a "circle" and calculated the "mass", "moments", and vector strength:

$$M = \int_{F_1}^{F_2} R(F)\,dF \tag{4}$$

$$P_x = \int_{F_1}^{F_2} R(F)\cos\left(\frac{2\pi F}{F_2 - F_1}\right)dF\,,\; P_y = \int_{F_1}^{F_2} R(F)\sin\left(\frac{2\pi F}{F_2 - F_1}\right)dF \tag{5}$$

$$V = \frac{\sqrt{p_x^2 + p_y^2}}{M} \tag{6}$$

Where R is the induced rate as a function of our frequency F ($F = f_0$ for *hm* maps and d_f for *tsh* maps), M is the "mass" (total area under the curve), P is the "moment", and V is the dimensionless vector strength of the tuning curve, or *tuning strength*. The range of valid vector strengths for non-negative curves is from 0 (identical firing rate across the map) to 1 (all the firing concentrated at a single point); negative induced rates could push the strength above 1. To evaluate this integral numerically we used the "trapezoidal integration" method (see Kaw A; http://www.mpia-hd.mpg.de/~mordasini/UKNUM/integration.pdf for a description of the method) on the induced rate. The limits of integration, F_1 and F_2, are 0 and f_0 for any *tsh* and $b_f/(k + 0.5)$ and $b_f/(k - 0.5)$ for any given *hm* and given order k. The results of a tuning strength calculation are illustrated in Figure 15.

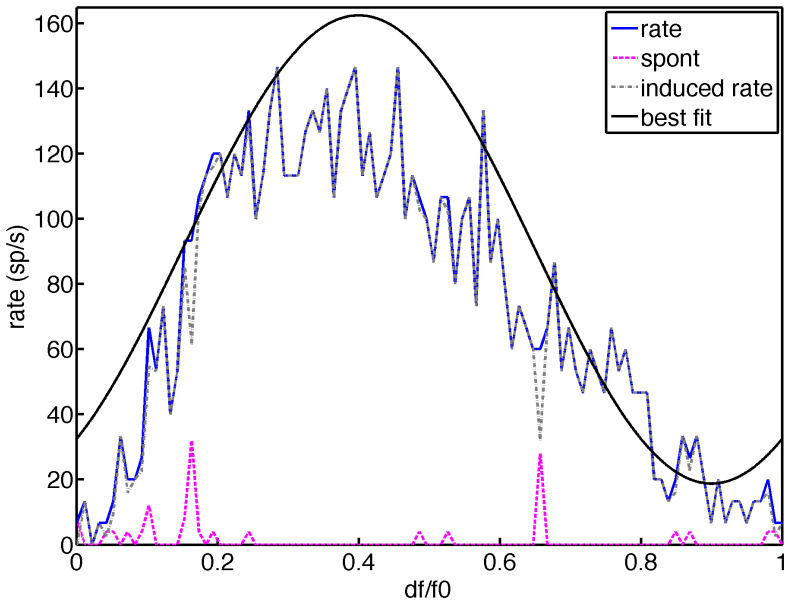

Figure 15: Calculating the tuning strength for the unit in Figure 14. The "induced rate" (spont subtracted from rate) is shown by the grey dashed line. The black line is the best-fit over all sinusoids that have one cycle as d_f varies from 0 to f_0. The best-fit amplitude was 144 sp/s and the best-fit offset (mid-level on the sinusoid) was 91 sp/s. The vector strength, which was 0.79 in our case, was half the amplitude:offset ratio of the sinusoid. The order (ratio of b_h to f_0) of 2.5 meant that we expected a peak at $d_f = 0.5f_0$ because that frequency put a component at b_h. The response curve itself peaked at about $d_f = 0.35f_0$ but the peak of the best-fit curve, at $d_f = 0.4f_0$, was closer to the expected value of $0.5f_0$. This discrepancy was due to the curve's non-sinusoidal shape. Since the spont was small in this example, the induced rate was almost the same as the rate, so the curves were almost on top of each-other.

Concatenating maps

Multiple *pure* and *hm* maps were concatenated if they were at the same sound level, as shown in Figure 16. Due to the complexity of the *hm* responses, many units needed multiple *hm* maps in order to control the resolution at various f_0s and/or get a close-up of the reference peak. To combine them, the frequencies (the *pure f's* or *hm* f_0*'s*) of all the tokens in each map group were concatenated into a single vector. The same was done with all of the rates and all of the sponts. The resulting freq, rate, and spont vectors were then sorted from low to high frequencies. This created a new map that had a larger number of tokens; the rates and sponts *weren't* averaged across tokens. For many units,

there were also multiple *pure* maps at the same sound level (30-40 dB above threshold), due to us testing for consistency. These were combined into a single *pure* map in the same way (not shown).

Two of the 22 units did not have an *hm* map near the 1st order peak (when f_0 is near the b_f). In those cases we used the *pure* as a surrogate for the *hm* for frequencies above the highest f_0 frequency taken. This is unlikely to be a large problem because both of the units had fairly sharp tuning. The first-order peak in the *hm*, when f_0 is near b_f, would have almost certainly been the same as the *pure* map's rate since neither of the two units would be affected by the components near $2b_f$,$3b_f$,$4b_f$, etc.

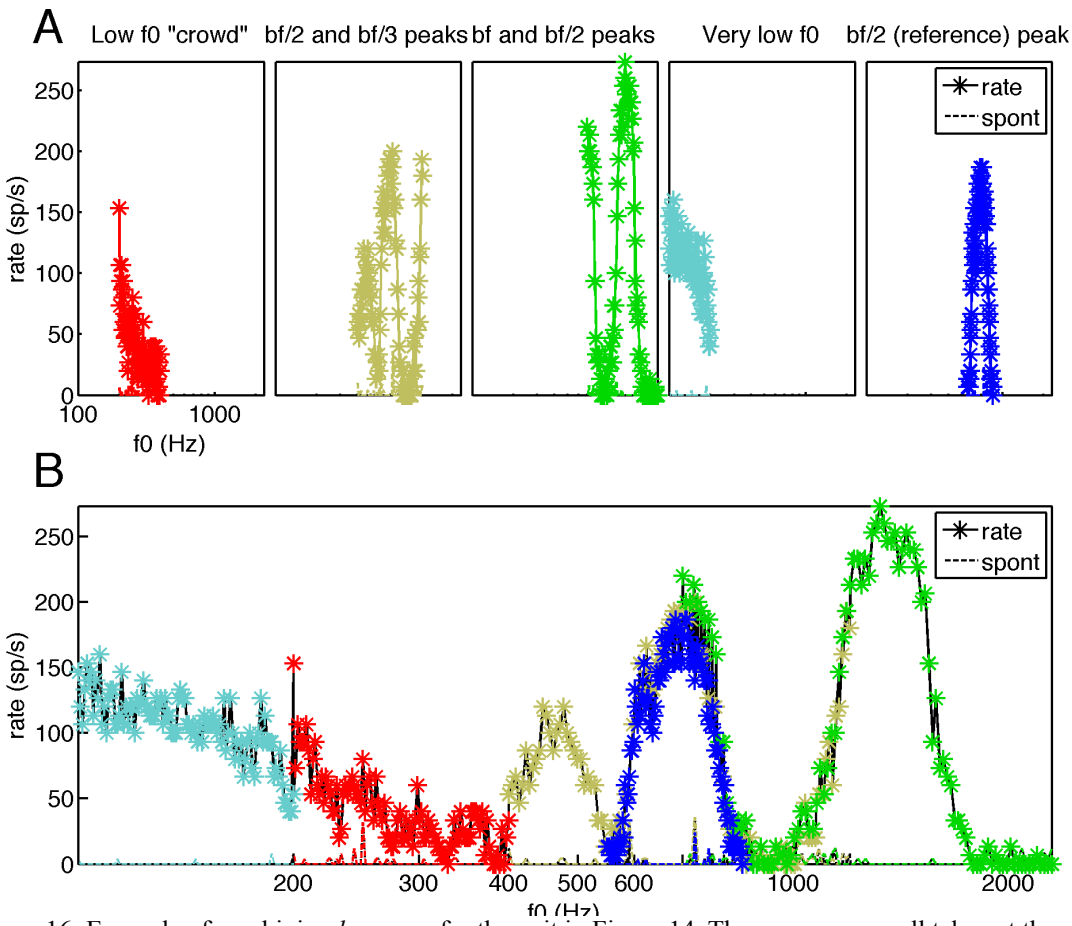

Figure 16: Example of combining *hm* maps for the unit in Figure 14. These maps were all taken at the same dB SPL. The frequency is on a log scale.
A, The individual *hm* maps, taken in chronological order. Each f_0 range was selected to investigate a different aspect of the response, as indicated in the title. All of the *hm*s completely covered the receptive field, so the unit probably responded in the same way had there been energy at $k\,f_0$ for all integers k up to "infinity". This let us combine the maps even though the higher f_0-range maps had less components (the maximum k was lower but the maximum frequency of the tokens, $k\,f_0$, were similar). The spont, being very small, is at the bottom of the plot.
B, The combined *hm* map. The black lines connect adjacent tokens in the combined map. Colored *'s correspond to tokens in the different plots from A. The maps sometimes overlapped, which increased the resolution in a given area. In particular, we took a detailed response at the reference peak (2nd order peak for this unit) which was used to find b_h.

Analyzing data on a per-order basis

Some of the results consisted of one datapoint per each order per each unit. For example, a unit could have *tsh* maps taken at 1st order ($f_0 = b_h$), "1.5'th" order ($f_0 = b_h/1.5$),

and 2nd order ($f_0 = b_h/2$), giving us three data-points. Except for units for which data collection was cut short due to lack of isolation, the maximum *tsh* order was determined manually based on the response flattening out (see Figure 17).

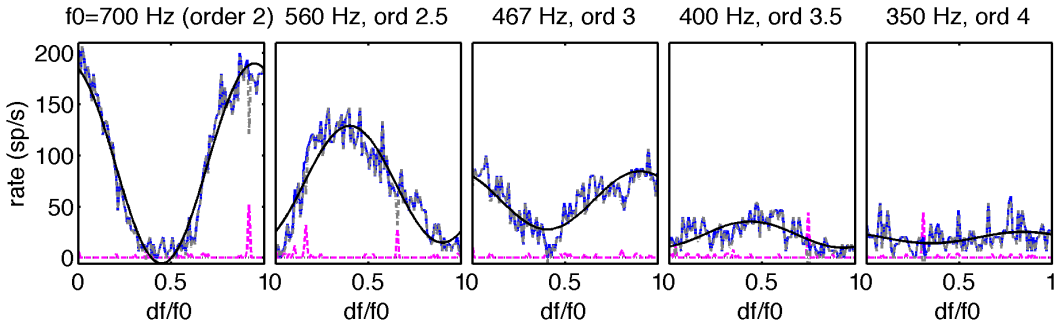

Figure 17: All the *tsh* data for the unit in Figure 14. All the half-integer orders from 2nd to 4th order are shown (there wasn't time to collect 1st and 1.5th order *tsh* data). Data are plotted in the same manner as Figure 15. *The 2.5th order data is the same data as in Figure 15*. At 4th order the response has flattened out; taking higher order *tsh* data would have been of little use. In this unit the overall response is much weaker to higher orders in addition to being flatter, most units' *tsh* data flattened out at high orders with little change in mean firing rate.

The available *hm* orders, on the other hand, were determined automatically. Isolation was not a constraint for the *hm* maps: we only included units in the data analysis for which at least one *tsh* taken, and *tsh* data were taken after the *hm* data. Instead, the number of peaks was determined using statistical significance, and each peak became one datapoint. The *hm* map was broken up into pieces given by $b_h/(k\pm0.5) < f_0 < b_h/(k\text{-}0.5)$, within this piece we expected a peak at b_h/k. The piece corresponding to $k = 1$ was *always* counted as a peak because the unit always had a b_f (rarely, the 1st order peak was not "significant" but the b_f was visible from the *pure* maps, which were not used in the significance test). Starting at $k = 2$, we calculated the statistical significance of each region's "non-flatness". Significance was defined as a vector strength larger than 95% of 10000 random shufflings (random permutations of the token frequencies). This process

stopped when the first non-significant region was found. Figure 18 shows the peak-detection results for three units.

For the data that was taken on a per-order basis, one dataset vector was made for each order by combining all units. However, not every order-unit combination worked out. For example, unit A may have had *tsh* data at orders 2, 2.5, and 3 while unit B had data at orders 2 and 2.5. In this case the 2nd order dataset included both units but the 3rd order only included A. At higher orders, fewer and fewer units had data available, so we limited data *presentation* to at most 6th order.

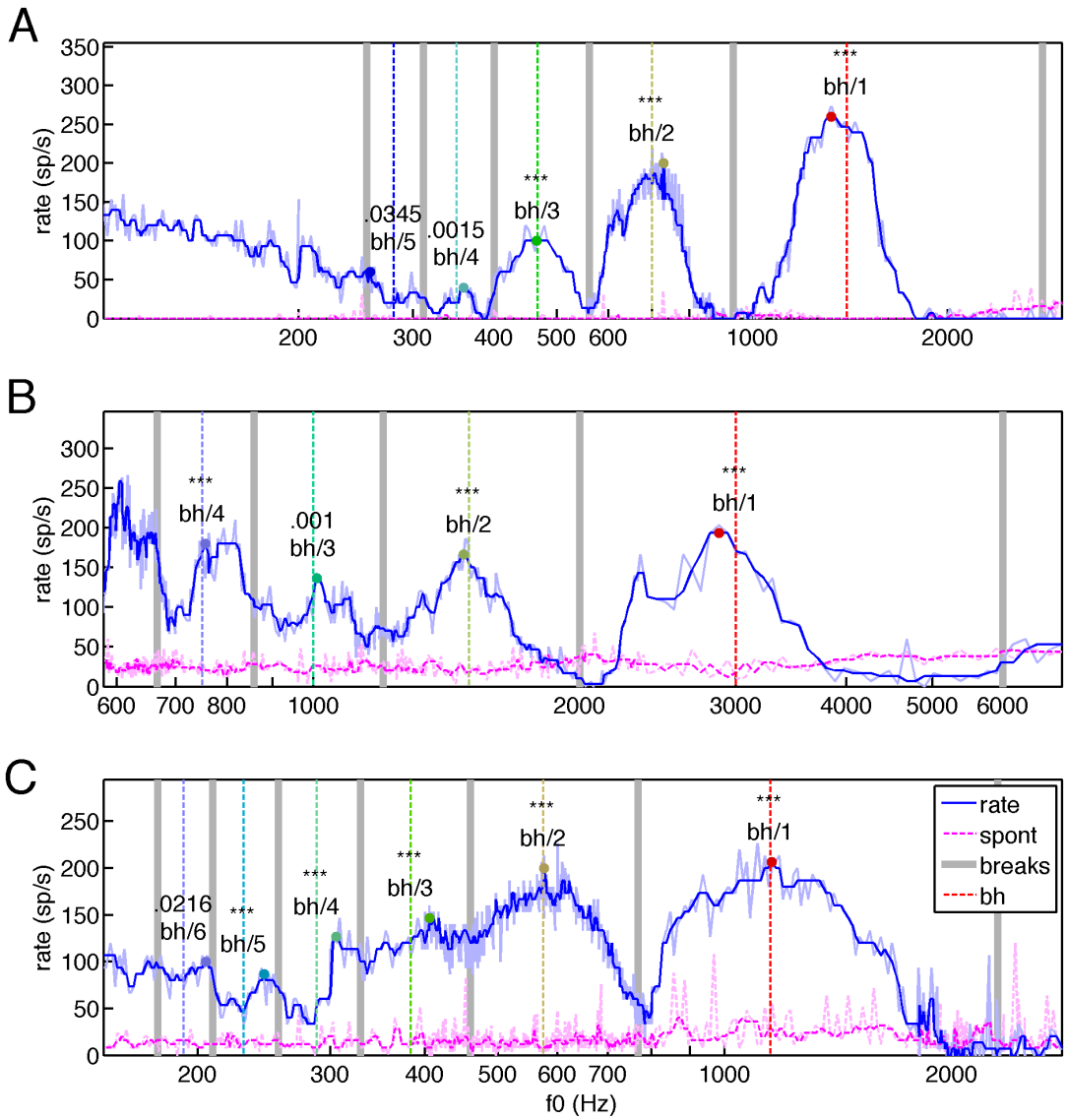

Figure 18: The (combined) *hm* map for three units and the results of the peak analysis. The faded colors are unsmoothed data and the stronger colors are smoothed data. There is a vertical dashed line at each expected peak location b_h/k, the one farthest to the right is at b_h. Each region is analyzed separately; solid grey vertical bars at $b_h/(k\pm0.5)$ denote the region boundaries (expected "valley" locations). The p-value of each region indicates the significance (*** means $p<0.001$) using the vector strength test described in the text above. The colored dots show the actual peak (frequency of maximum smoothed induced rate) locations.
A, The analysis on the unit in Figure 14 found 5 significant orders. For orders 1-4 the expected and actual peak and valley locations coincided. However, for 5th order there was not a nice agreement.
B, Another unit. The peaks were very close to the expected location for all 4 orders and the valleys were reasonably close to the boundaries. This was one of only 2 units that had a significantly incomplete *hm* map (another 6 units had *hm* maps that *may* be slightly incomplete). This *hm* map was taken under the older paradigm with 7 components only. Despite the limitation, we already are seeing the response map start to level out as indicated by the non-zero valley floors.
C, A third unit. Peaks have been found for 6 orders but only the first two fit nicely with the expected locations. The other peaks are phase-shifted to higher frequencies compared to the expected locations.

Linear additive model

We measured non-linear effects in a neuron by comparing the neuron's response rate in a *hm* or *tsh* map to that of a linear additive model:

$$R = \sum_k \left(R_{ind}(f_k) \right) + R_s \tag{7}$$

Where R is model's rate, R_{ind} is the smoothed induced *pure* rate (which is a function of f_k, the k'th frequency component of the sound), and R_s is the *pure* spont rate for our unit. By subtracting the spont and then adding it back in, our model allowed a little inhibition because the rate sometimes went below the spont (negative induced rate) in the sidebands. For numerical stability, we only considered frequencies between 1/3 and 3 b_f. Even with this restriction, f_k was often beyond the limits of the map. In these cases we used the median of the spont for the lowest 3 and highest 3 tokens of the *pure* map. The rate for the *pure* response in out-of-bounds cases should equal the spont (since the range of the map was designed to cover the entire receptive field); we used the "edge spont" instead of the overall spont since it was less likely to be affected by sound.

The individual *contributions*, $R_{ind}(f_k)$, can be visualized as curves superimposed on the *hm* or *tsh* map. All maps are parameterized by a frequency parameter F, which is f for a *pure*, f_0 for an *hm*, and d_f for a *tsh*. The k'th contribution curve is generated by plotting $R_p(f_k(F))$ over all F values. Each contribution curve is shifted and scaled *opposite* to how the components are related to F. For an *hm*, $f_k(F) = f_k(f_0) = kf_0$. The curve plotted, $R_p(k f_0)$, is identical to the *pure* rate $R_p(f)$ except that it has been *shrunken* by a factor of k on the frequency axis. Similarly, the *tsh* curve copies of the *pure* rate shifted *downward* on the frequency axis by $(k\text{-}1)f_0$. Figure 19 shows a visualization of the contributions.

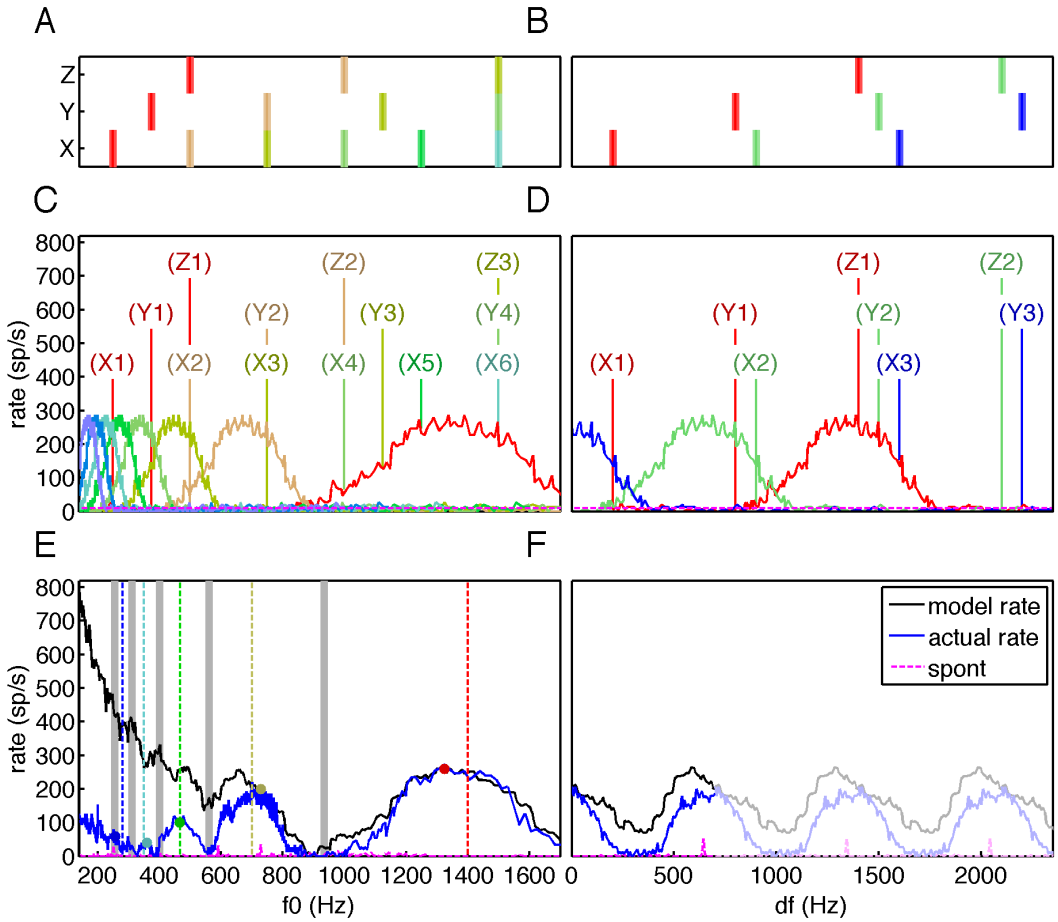

Figure 19: The linear additive model for an *hm* and *tsh* map for the unit in Figure 14.
A,B: schematic of the stimuli for the *hm* and the *tsh*, respectively. Three example tokens, labelled "X", "Y", and "Z", are shown.
C, D: The pure contributions to the linear model's response as a function of f_0 and d_f for the *hm* and *tsh*, respectively. The colors have a 1:1 relationship to those in in A and B and the red curve all the way to the right is the unmodified *pure* response. However, the colors are "backwards" since the formula for the contributions is inverted from that of the components. The frequency of each component for each token is labelled. The mean spont is shown as the dashed line near the bottom (the per-token spont was used for the calculations but the per-token spont was close to the mean spont in most cases).
E, F: The simulated (linear model) rate plotted with the actual rate for the *hm* map in A and the *tsh* map in B, respectively. The axes scales for E and F are the same as for A and B, respectively. E also shows the harmonic peak regions as in Figure 18. In E the model's prediction tends to infinity at low f_0s because more and more energy ends up in the receptive field, this is visualized as the contributions "bunching up" in C. The actual rate, on the other hand, shows saturation. The *tsh* in F covers a single cycle ($d_f = 0$ to f_0), shown as the darker portion of each curve. The faded curves, exact copies of the map that have been shifted right by multiples of f_0, show the hypothetical responses for *tsh* tokens with $d_f > f_0$ under the assumption that the stimulus is perfectly "circular".

Building a harmonic template unit out of pseudopopulations

A linear additive model was also used to construct a harmonic template unit that was designed to be receptive to harmonics of a particular f_0 and be sensitive to mistunings, as were the template units in Feng (2013: p34-35). Instead of breaking up a token into individual components and summing the spiking rates to each component, we presented the sound to a bank of *pseudounits* and calculated a weighted sum of the responses.

Each *pseudounit,* generated using the method in May et al. (1998) and Cai et al. (2009), was identical to our IC unit but "scaled" to have a different b_f. Each map token was represented as a vector of frequencies $\boldsymbol{s}$. A *tsh* token with f_0 = 500Hz and d_f = 100Hz, for example, would have $\boldsymbol{s}$ = {600, 1100, 1600, …} Hz. This simplification disregarded time-domain effects such as envelope modulation and phase. The response of the pseudounit to a sound **s** is defined as:

$$R'(\boldsymbol{s}) = R(\frac{b_f}{b'_f}\boldsymbol{s}) \tag{8}$$

Where b_f and b'_f is the best frequency of our unit and the pseudo-unit, respectively. A *higher* b'_f pseudounit will "hear" each component of **s** as being of a *lower* frequency, thus the division by b'_f in the equation.

The weighted sum of the pseudopopulation responses gave the input *current* to the harmonic template model unit. The model unit received excitatory inputs with $b_f = k_e b_{f0}$ and inhibitory inputs with bf $b_f = (k_i + 0.5)b_{f0}$, for integer k_i and k_e and a given "preferred fundamental frequency" b_{f0}. An *hm* token with $f_0 = b_{f0}$ is the key that fits into the spectral

lock: it lands on the b_fs of all the excitatory inputs but none of the inhibitory inputs. This token should yield a higher input current than either *hm* tokens with a different f_0 or non-harmonic stimuli such as mistuned complexes and tones. A realistic template model would include non-linear effects such as thresholds and saturation. However, accounting for this would require specifying more degrees of freedom, obfuscating the underlying results. For the data presentation, we showed the (linear) net input instead.

The input to the template neuron can be expressed in terms of the responses of the IC unit to a variety of sound stimuli:

$$I(\boldsymbol{s}) = \sum_{k=m}^{M} R\left(\frac{b_f}{k\, b_{f0}}\boldsymbol{s}\right) - \sum_{k=n}^{N} R\left(\frac{b_f}{(k+0.5)\, b_{f0}}\boldsymbol{s}\right) \tag{9}$$

Where I is the input to the template neuron when presented with sound $\boldsymbol{s}$. The values of m, M, n and N determine which orders feed into our template.

The choice of b_{f0} is arbitrary since I is unchanged if both b_{f0} and $\boldsymbol{s}$ are scaled by the same amount. For convenience, we defined $b_{f0} = b_h/k_{ref}$, (b_h is the "harmonic best frequency" and k_{ref} is the reference order we used to calculate b_h). This put the b'_fs at similar values to our b_h and b_f. Figure 20 demonstrates the calculation of the template response using this convention. Other methods for setting b_{f0} are possible as well.

Regardless of our conventions, in order to test mistunings of a harmonic complex at b_{f0}, the *tsh* frequencies had to be integer and half-integer *fractions* of b_h. For *pure* or *hm* maps, scaling the $\boldsymbol{s}$ vector can be done by varying f or f_0 (moving along the map). However, for *tsh* maps the *spacing* d_f also needs to be scaled. To get the template's response to mistuned complexes, i.e. *tsh* tokens with $f_0 = b_{f0}$ and $d_f > 0$, the simulated *tsh* must have a spacing $f_0^{sim} = b_{f0}$. Each pseudounit will demand a different real sound stimuli

$f_{0,k}$ and a range of d_f values from 0 to $f_{0,k}$. By our convention, the b_fs of each *pseudounit* are at $b_{f0}k$ and at $b_{f0}(k+1/2)$, for integer k. The real sound frequencies are given by $f_0 = b_h/(b_{f0}k)b_{f0} = b_h k_{ref}/(b_h k)b_h/k_{ref} = b_h/k$ and $f_0 = b_h/(k+0.5)$. As expected, the convention of choosing b_{f0} has no effect on the requirement that the real *tsh* maps f_0 values must be at integer and half-integer *fractions* of b_h. Thus the choice of which *tsh* maps were collected influenced which pseudounits could be included in the input to the template; taking *tsh* data (when possible) all the way out to the point at which the *tsh* became mostly flat (see Figure 17) made it less likely that no "important" (non-flat response) pseudounits were missed. The requirement for taking data at particular *tsh* f_0s meant that we could not model multiple templates with different b_{f0}s.

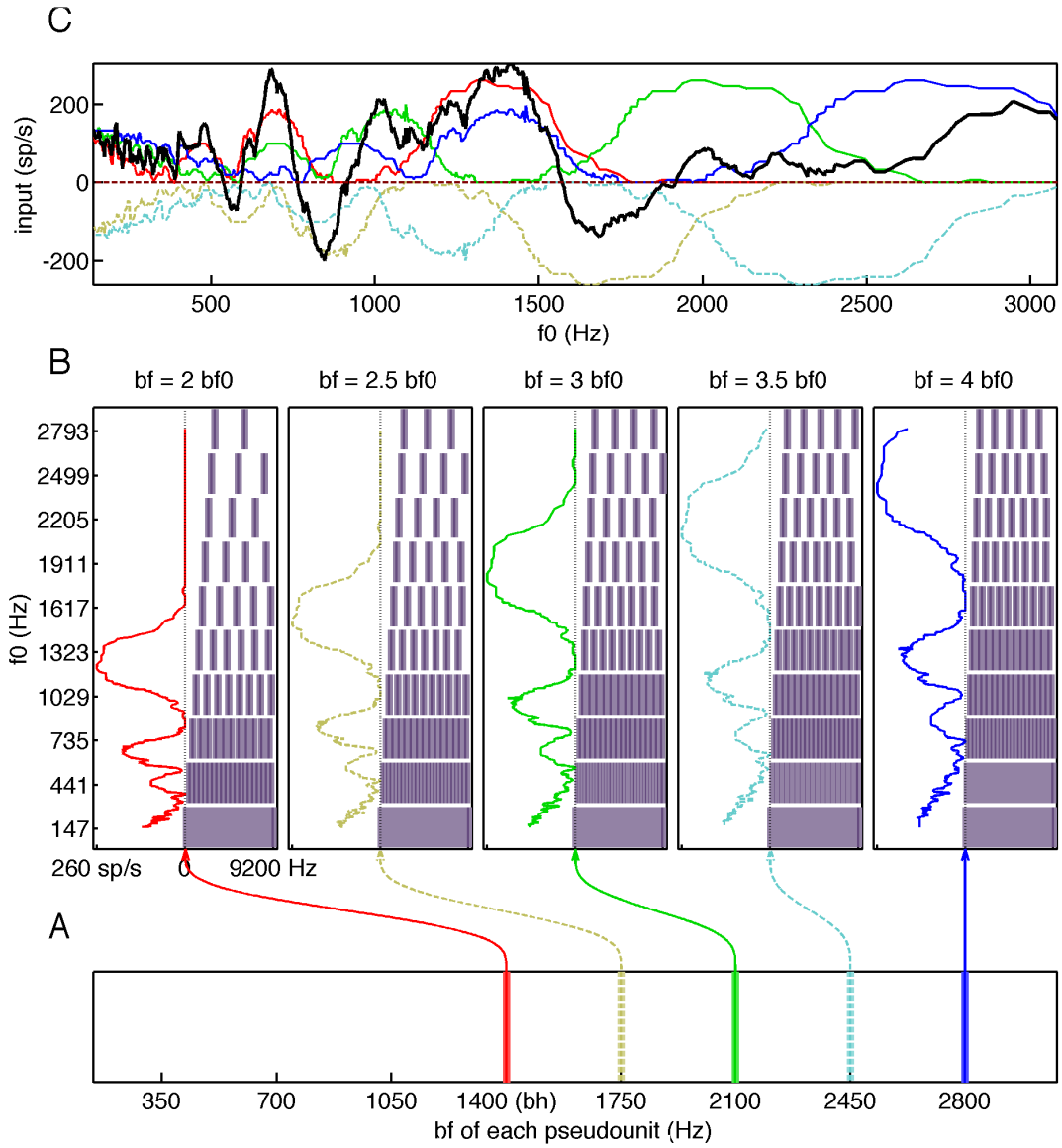

Figure 20: How to calculate the input current to a model harmonic template neuron, for an *hm* map, from the responses of five pseudounits (using the unit in Figure 14) which converge on the template unit.
A, Locations of pseudo-units, color-coded from red (low frequency) to violet (high frequency). The $b_{f\text{pseudo}}$s are at integer and half integer multiples of 1/2 of the real b_h. Each bar is the b_f of a pseudounit. Dashed lines are inhibitory pseudo units, solid are excitatory. In this unit the bar farthest to the left corresponded to the real unit's b_h. Had there been more time we would have taken data for $b_{f\text{pseudo}}$s less than b_h as well, however.
B, The response of each pseudounit to the *hm* map. The curve on the left is the firing rate, while the bars on right are the real stimuli schematics (with the color removed) of tokens at values of f_0 presented to the template. Different subplots correspond to different $b_{f\text{pseudo}}$s. The axes scales are the same for all the subplots in B. The curves are different, however, because the higher $b_{f\text{pseudo}}$s were presented sounds with lower real f_0s.
C, The input to the template unit as a function of the f_0 presented to the template. The colored lines are the inputs of each pseudounit (inhibitory inputs are dashed and have been multiplied by -1), while the black line is the total response.

Cat Auditory-nerve model

Simulated cat AN units were fed through a similar data processing pipeline so that the IC's spectral processing could be compared to that of the AN. The cat auditory nerve model from Zilany et al. (2009) was used.

The Inner Hair Cell (IHC) output served as a proxy to the model since the AN spike rate calculation involves a computationally expensive Monte-Carlo simulation of the IHC/AN fiber synapse. Both the IHC output and the spiking rate had similar properties in response to harmonic sounds as the units in Cedolin and Delgutte (2010). The model allowed us to select the unit's b_f and chose between three options of fiber: “low”, “medium”, and “high” spontaneous rates (we always used the “medium rate”). Next we specified the waveform and sampling rate to stimulate the nerve fiber; we used the same parameters as we did for the real stimuli (including a 10ms ramp with a 150 ms total stimulus duration).

The cat model required adjusting the dynamic range. The AN units, in a model based on the anesthetized chinchilla, typically had a dynamic range of 30dB to tones (Yates, 1990). Although a few high-threshold (about 40 dB SPL) units had a much wider dynamic range (Yates, 1990), we were interested in capturing the behavior of the more moderate 30dB-range units. However, our cat model, even for moderate (about 20 dB SPL) threshold units, behaved like the high dynamic-range units in Yates (1990) (see Figure 21A, red dashed curve). To fix the dynamic range problem we normalized the maximum output of each neuron to 1 and presented all inputs as dB relative to threshold (the sound level that yielded 5% of maximum firing rate for tones at b_f, over all sound

levels up to a limit of 100dB SPL). We then passed the rate through a logistic saturation step so that the output was always 95% of the maximum output 30dB above threshold (the rate for the token as a whole behaved differently since the logistic saturation step was done *before* time-averaging). This saturation step added a small amount of spont to the model units (see Figure 21A, purple and blue curve). More importantly, the step made the *RLV* curve look more like the typical units in Yates (1990). The AN model's response is summarized in Figure 21.

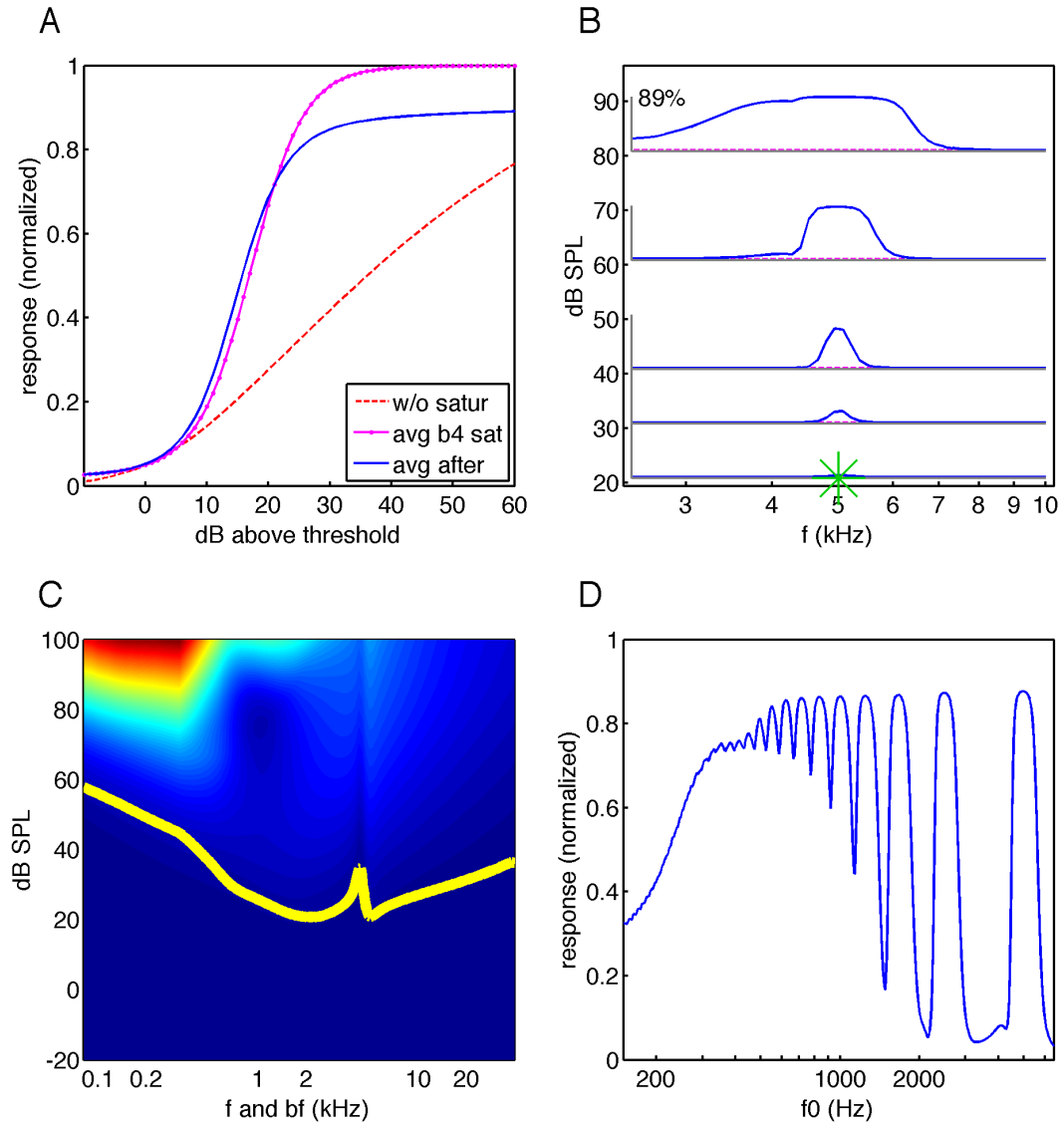

Figure 21: Responses of the cat AN model.
A, Normalized response to an RLV map for a tone at b_f for a mid-frequency (5kHz) unit. The dynamic range without saturation was >60dB. Averaging the rate during the sound stimulus first then applying saturation produced, as designed, a 30dB dynamic range curve. However, for the calculations we applied saturation *first* and then averaged (blue curve); this curve had a slightly lower asymptotic rate.
B, Response maps as in Figure 14 for the model unit, the normalized response peaked at 0.89.
C, Heat map of response to tones for units with b_f matched to the stimulus frequency, before applying normalization and saturation. This is *not* a response map because we are varying b_f along with the tone's frequency as well. The yellow line is the threshold, the dB SPL that elicits a response 5% that of the maximum response of the given column (we limited sound level to 100dB SPL for determining the "maximum"). The response to the lower frequency units was much higher and there was a spectral notch at 4.3 kHz superimposed on an overall V-shaped audiogram. Normalization to dB with respect to threshold and to maximum firing rate allowed us to compare units with different b_fs.
D, Response of the 5kHz unit to an *hm* map with each component 40dB above threshold. 5 peaks are strongly "resolved" and at least 10 peaks are visible. At very high harmonic number the response drops off, presumably because the frequency sweeps pass through b_f less and less often as f_0 is lowered.

Results

Tuning properties

A total of 22 IC units, which were taken with the up-to-date protocol and had at at least one *tsh* data-point, were analyzed. The tuning properties of each unit and the response maps of a few examples are given in Figure 22. The units had a wide range of b_fs, thresholds, and Q_{40}s. The lowest thresholds at low frequencies were close to the behavioral audiogram in Osmanski (2011) but at higher frequencies were about 15 dB *lower*. There was a weak correlation between the log of b_f and Q_{40} (linear regression: adjusted $r^2 = 0.22$, $p = 0.015$) with a notable absence of low b_f -high Q_{40} units. The use of b_h to approximate b_f worked well, evident in the good correlation between the two ($r^2=0.98$ on a log-log plot with a 1:1 linear model) shown in Figure 23. The one outlier had a poorly defined b_f (see Figure 22).

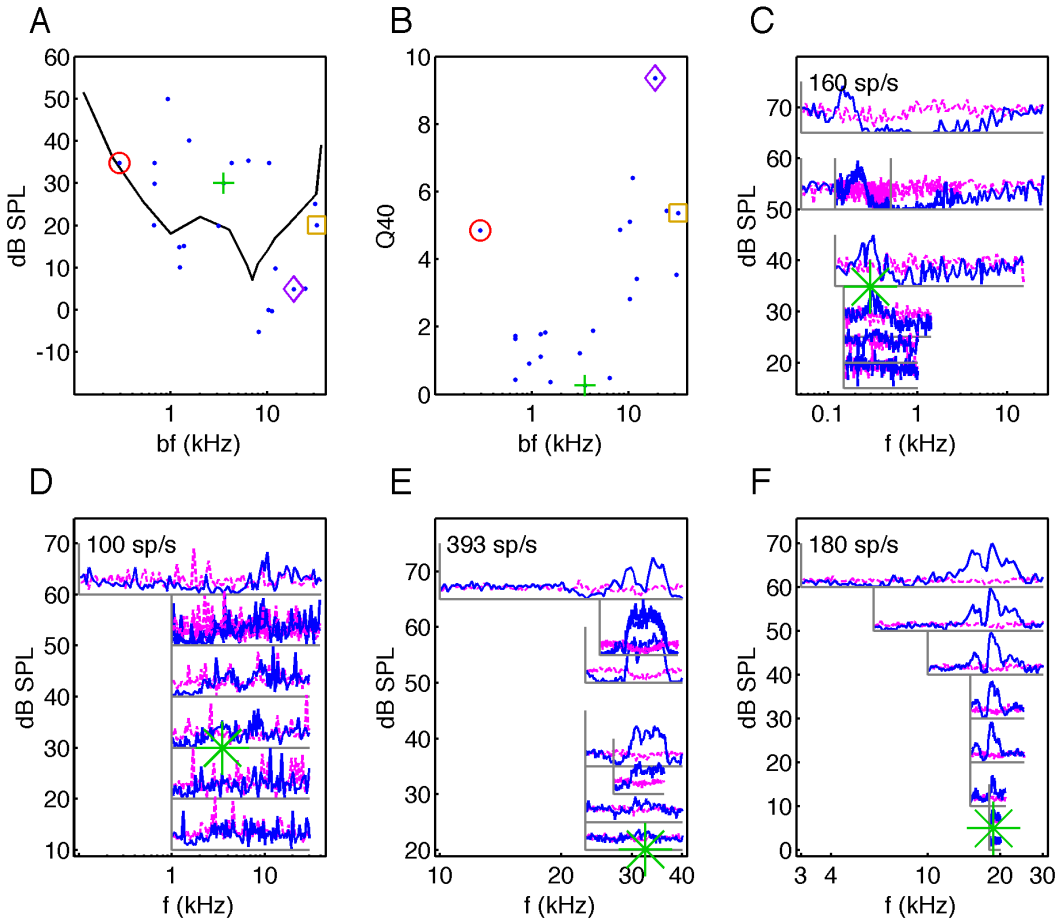

Figure 22: Tuning properties of our 22 units.

A, The b_f against the threshold dB SPL and Q_{40}, colored dots correspond to examples shown in C-F. The marmoset audiogram (behavioral threshold) from Osmanski (2011) is the black line. Some units, to high frequencies in particular, exceeded the boundaries of the audiogram (had significantly lower thresholds). The peak sensitivity near 10kHz corresponds to the resonantly amplified frequencies of the marmoset external ear (Slee and Young, 2009).

B, The b_f vs Q_{40} for each unit.

C-F: Figure 14-style response maps.

C (corresponds to the o symbol), The unit with the lowest b_f (0.29kHz). This was a type-O unit with a strong, wide inhibition at frequencies above b_f.

D (+ symbol), The lowest Q_{40} unit (0.275). The Q_{40} value is very low because the peak at the 60dB level (there was no 70dB map) was at a much higher frequency than the b_f. Although the peaks were questionable, there *is* a clear pattern of inhibition around 100Hz, apparent in all the *pure* maps. This response map also contains several artifacts that were caused by animal-generated sounds (random sharp peaks); this was our most poorly resolved response map.

E (□ symbol), The unit with the highest b_f (32kHz). There was inhibition up to at least 40kHz, which was beyond the the marmoset's hearing range reported in Osmanski (2011).

F (◇ symbol), Highest Q_{40} unit (18.7). This unit had a very sharp three lobed tuning pattern, and the Q_{40} was so high in part because only the center lobe reached above 50% of the maximum.

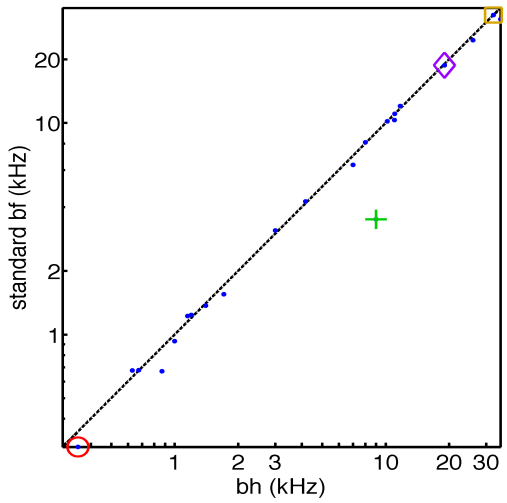

Figure 23: The relationship between b_h and b_f plotted on a log scale. The dashed 1:1 line fit the data quite well. The highlighted symbols are for the same units as in Figure 22. The median difference between b_f and b_h was 4.3%.

Responses of ICC units to tones, harmonics, and mistuned complexes

None of the 22 units showed harmonic selectivity according to the criteria in (Feng, 2013: p71). A "template unit" must satisfy two quantitative thresholds which we calculated for each *hm* and/or *tsh* order (see Analyzing data on a per-order basis). Firstly, there must be more response to harmonics than tones. This ratio is quantified in Feng (2013: p11) using the facilitation index:

$$F_{I,k} = \frac{R_{max,k}(hm) - R_{max}(pure)}{R_{max,k}(hm) + R_{max}(pure)} \quad (10)$$

Where $R_{max,k}(hm)$ is the height of the smoothed k'th order *hm* peak and $R_{max}(pure)$ is the height of the smoothed *pure* peak. An F_I of below 1/3 excluded being a harmonic template unit (Feng, 2013: p71). This cutoff corresponded to twice as much maximal response to harmonics as to tones. Unlike Feng (2013), we calculated $F_{I,k}$ for each order

with a significant *hm* peak instead of the whole unit.

The other criterion to be a template units is sensitivity to tuned vs mistuned complexes. We defined a periodicity index that is similar to the one in (Feng,2013: p13):

$$P_{I,k}=\frac{R_{max,k}(tsh)-R_{min,k}(tsh)}{R_{max,k}(tsh)+R_{min,k}(tsh)} \tag{11}$$

Where $R_{min,k}(tsh)$ and $R_{max,k}(tsh)$ is the minimum and maximum height of the smoothed induced rate for the *k*'th order *tsh* map, respectively. A P_I of at least 0.5 was necessary for a unit to be considered a template unit in (Feng, 2013: p71). Our P_I definition always yielded *higher* results than the one in Feng (2013) because it compared the maximum to the minimum firing rate instead of the $d_f = 0$ to $d_f = f_0/2$ firing rate.

Most of the units had weaker maximal responses to the *hm* than their *pure* maps. This was indicated by the $F_{I,k}$ values (used on the smoothed induced rate) being mostly negative. This result was statistically significant. If positive values were as likely as negative values, the number of positive values *h* would be a sample from a binomial distribution with probability 0.5 and with the number of trials *n* equal to the total number of samples. We can assign a p-value to testing for more (-) outcomes than (+) outcomes: p is the cumulative probability of our (*n*,0.5) binomial distribution between 0 to *h*, inclusive. By this binomial test we have $p<0.05$ for k=1,2,3 and 5 and $p<<0.001$ combined over all *k*. The relative responses to the *pure* and *hm* maps are summarized in Figure 24.

The peaks of the *tsh* maps showed a similar trend of weaker response. These maps acted like close-ups of each of the *hm* peaks. When d_f was 0 they tested the response to a harmonic complex with a tone at b_h (where we expected the peak to be). Even if the value

of b_h was slightly different from the "true" b_f, there was a non-zero value of d_f (which was small so the stimulus was nearly harmonic) which placed a component at the "true" b_f. The main advantage of the *tsh* data is that they are at a much high-resolution since 100 tokens were taken across a single peak region. Indeed, the *tsh* provide confirmation of the preference against *hm*: although they had a slightly *higher* peak response for the first-order peak (binomial test, $p = 0.03$), for orders 3,4,5 the peak was *lower* ($p<0.05$), as it was for the overall the trend ($p = 0.00015$). Figure 25 summarizes the *tsh* data.

Some units had a *much* weaker response to *hm* than *pure* maps. The unit in Figure 26 had the lowest (most negative) F_1; it's response to harmonics was suppressed well below the spont. The unit with the most *absolute* reduction in firing rate (most negative F_1 when the denominator of eq. 12 is set to 1), shown in Figure 27, displayed a different pattern. This unit had progressively lower peaks to harmonics of increasing order followed by a "build-up" pattern where the response to the *hm* partially recovers at low f_0 (but without a distinct peak structure).

Most units showed little difference between *hm* and *pure* response peak height. These units typically had single, sharply-defined peak to *pure* and multiple peaks to *hm* maps (near b_f/k for integer k). At higher-orders, the *hm* map tended to leveled out to an intermediate value. Figure 28 shows an example unit that is only slightly selective to *hm* maps; most units with a small amount of selectivity for or against harmonics showed a similar pattern.

Although two units were strongly selective for *hm* stimuli over *pure* stimuli, none of our units were harmonic template units. A necessary condition to be a template unit, as

defined in Feng (2013: p71), is to have an $F_{I,k}$ >1/3 *and* a $P_{I,k}$ > ½ for at least one order k. The highest maximum $F_{I,k}$, over all units and k values, was exactly 1/3 (this value is exact was because the number of spikes has to be integral) but the P_I (using our metric that *overestimates* P_I) for the only *tsh* map for this unit was just 0.26. This *tsh* f_0 was near the best f_0 for the *hm*, as were the *tsh*s in Feng (2013). Also, the *hm* map had a smooth, broad peak (see Figure 28), leading to the conclusion that this was nothing like the (sharply-tuned) template units in Feng (2013). The next highest maximum F_I, 0.29 (with an average of 0.25 across the unit), belonged to a unit with similar properties: broadly tuned *hm* and *tsh* responses with P_I below 0.5 for *tsh* orders >1. The third highest maximum F_I was 0.17 (average of 0.04), this unit is shown in Figure 28. It did not have a broad tuning with a generic preference for multi-component stimuli, instead it was typical of the units with $F_I \approx 0$ and reasonably sharp tuning.

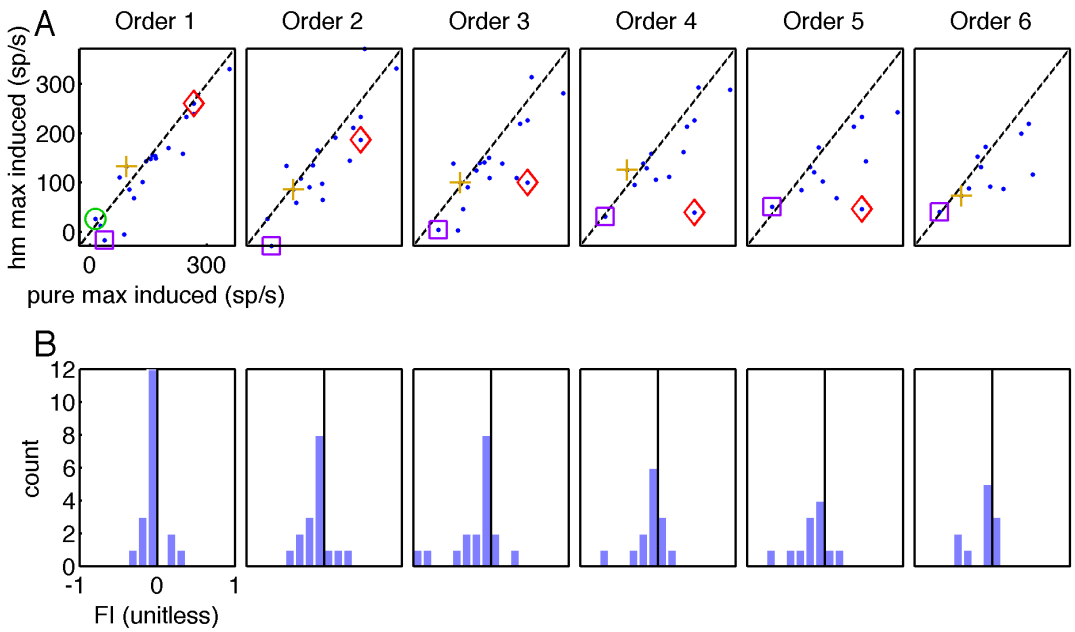

Figure 24: Facilitation indexes.
A, The peak firing rate for *pure* vs *hm* maps for the first 6 orders (the *pure* peak is identical for a given unit for each order but the *hm* peak changes). Colored symbols indicate examples shown in Figure 26-Figure 29.
B, Facilitation index histograms for each order. The vertical bar dividing each histogram between left and right is where $F_I = 0$.

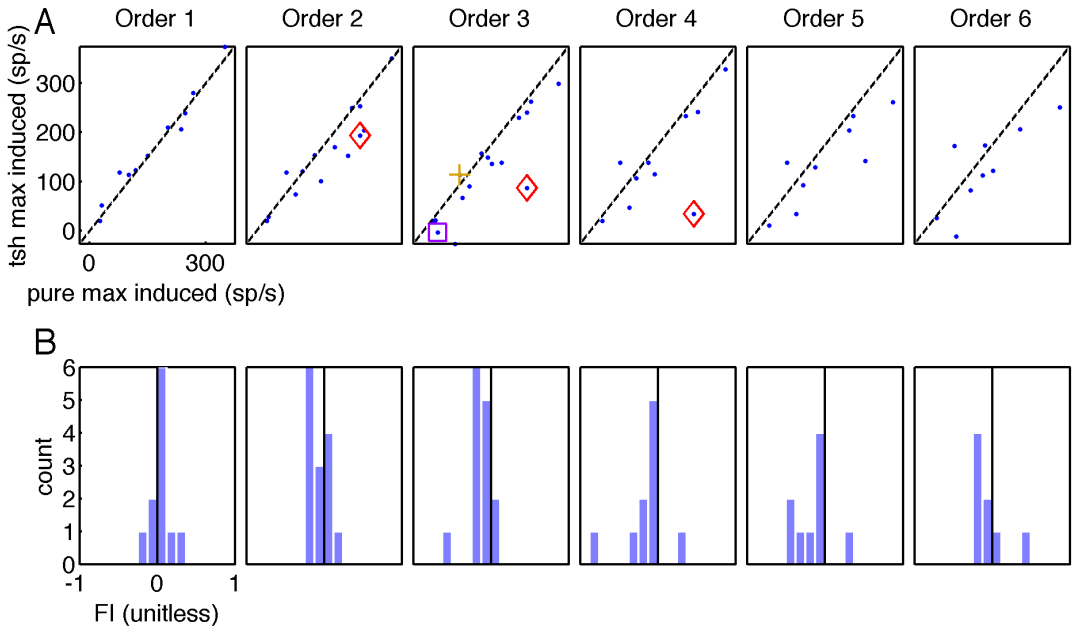

Figure 25: Same as Figure 24 except that we are using the *tsh*s in place of the *hms*. The *tsh*s provide a more robust peak estimate because they act as high resolution closeups of each peak, but are a smaller dataset.

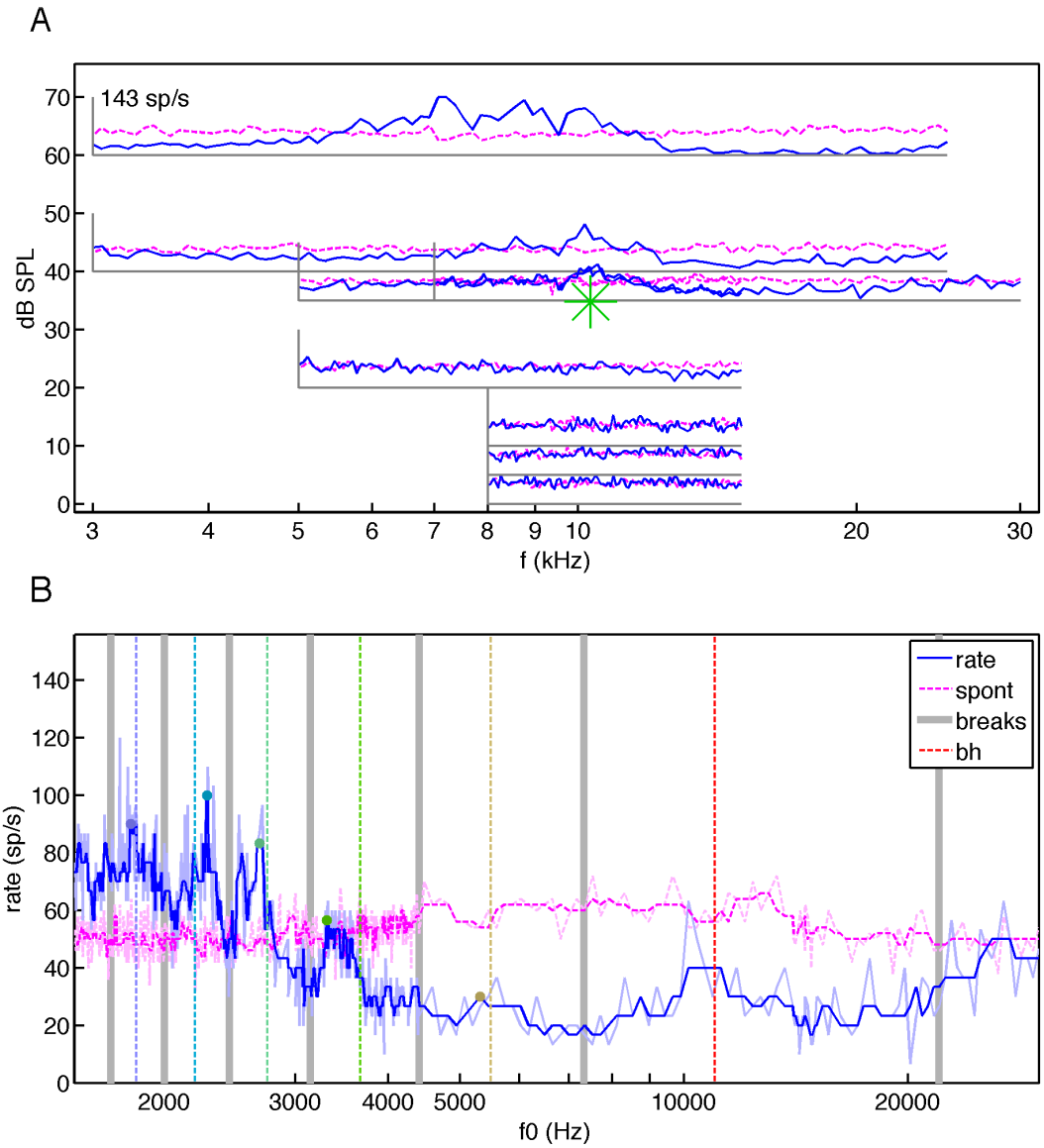

Figure 26 (corresponds to the □ symbol in Figure 24-Figure 25): The unit with the minimum facilitation index. The averaged F_1 was *negative* 1.6 due to the below-spont rates for the *hm* near b_f (the F_1 values were as low as -6.5 at low orders because the induced *hm* rates were almost as negative as the induced *pure* rates were positive, making the denominator of eq (10) close to zero).
A, The response map, as in Figure 14, for this unit.
B, the harmonic response as in Figure 18.

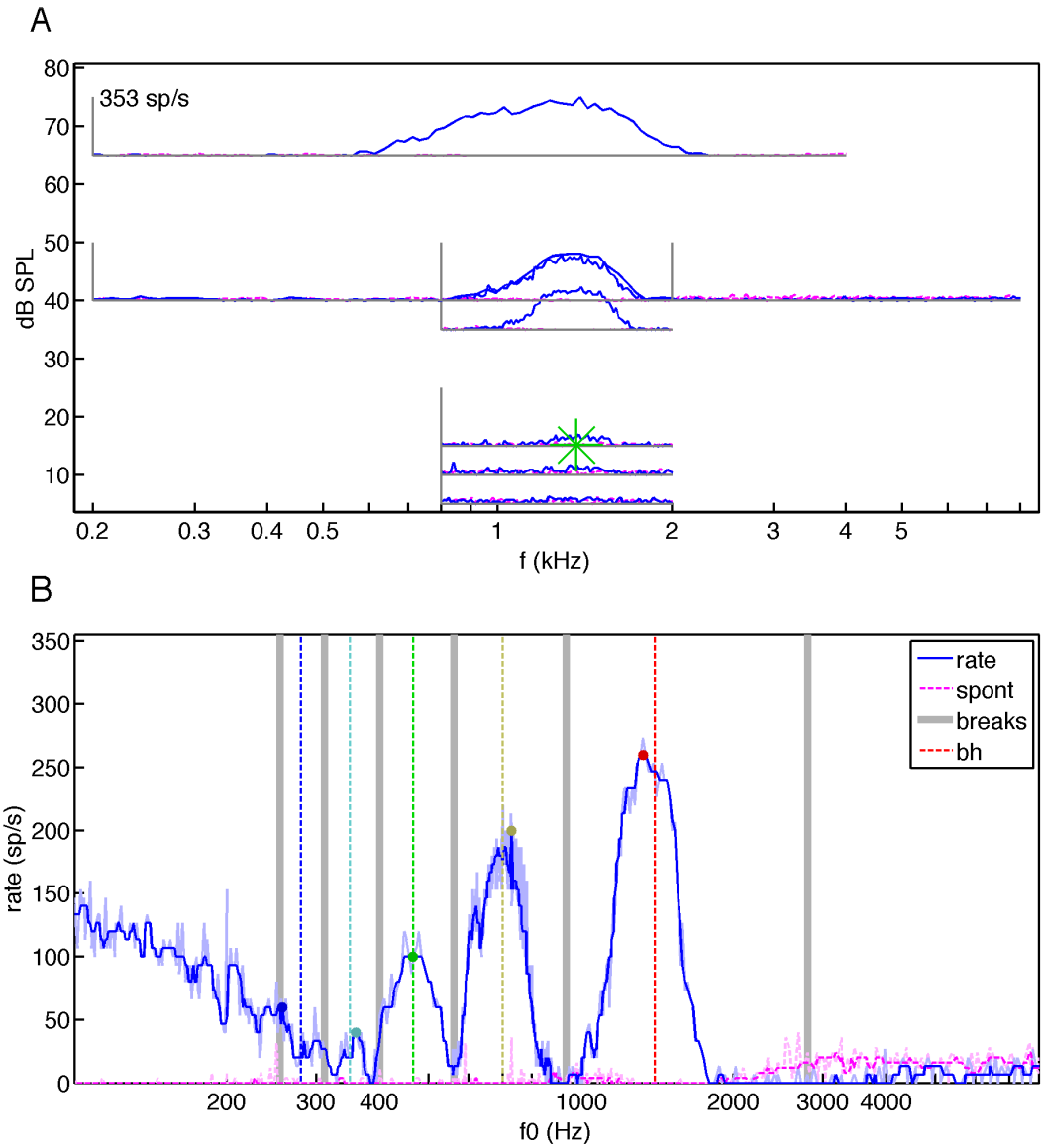

Figure 27: (◇ symbol in Figure 24-Figure 25): The same as Figure 26 but for the unit with the minimum "absolute facilitation" (137 spikes/s more response to *pure* than *hm* maps averaged over all orders). The higher order *hm* peaks show more and more inhibition, up to order 4, after which there is a gradual *increase* in the firing rate with decreasing f_0. *This unit is the same unit as in Figure 14*.

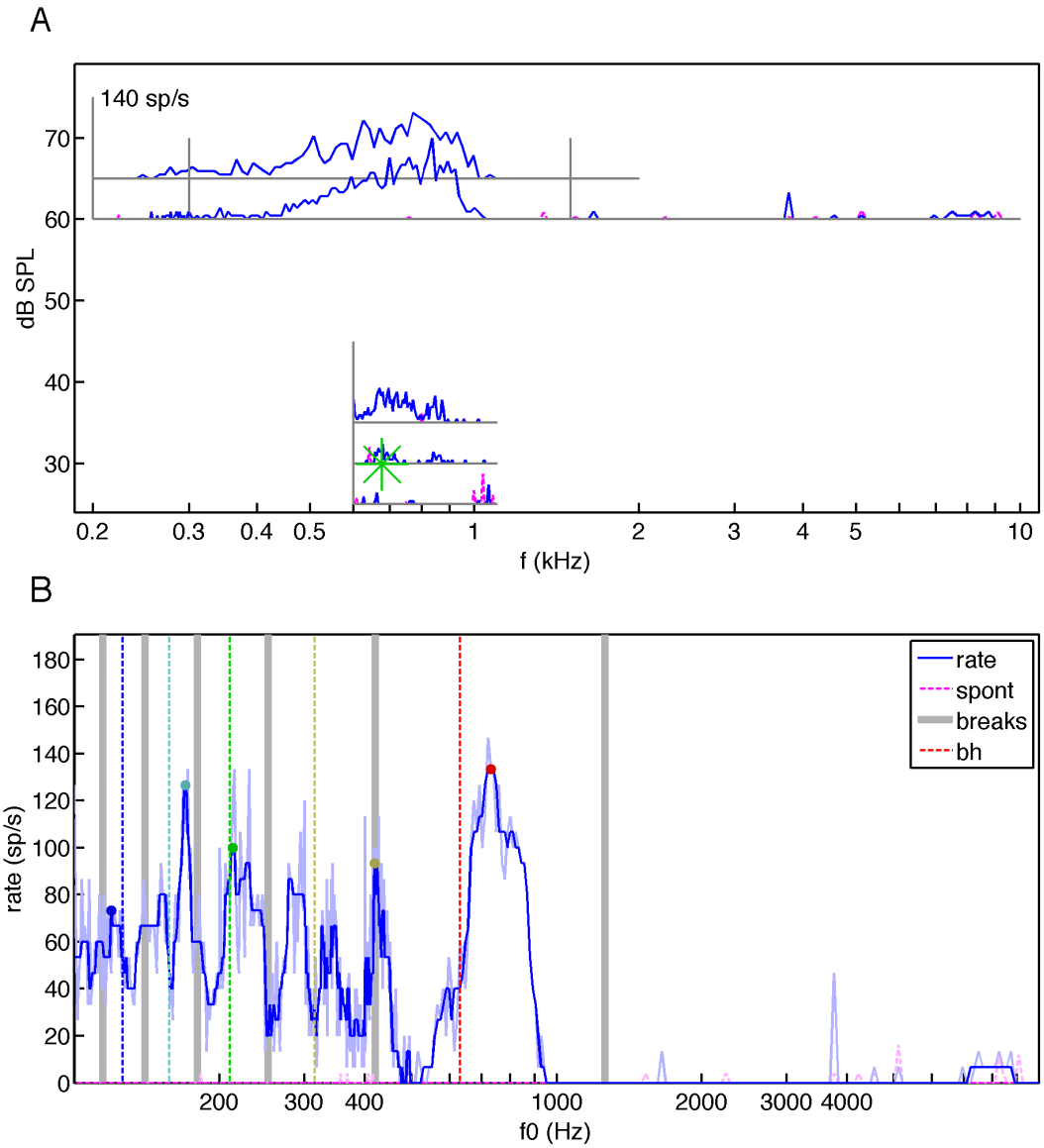

Figure 28: (+ symbol in Figure 24-Figure 25): Like Figure 26 but for a unit that was barely selective for harmonics (an average F_1 of 0.04, maximum of 0.17).

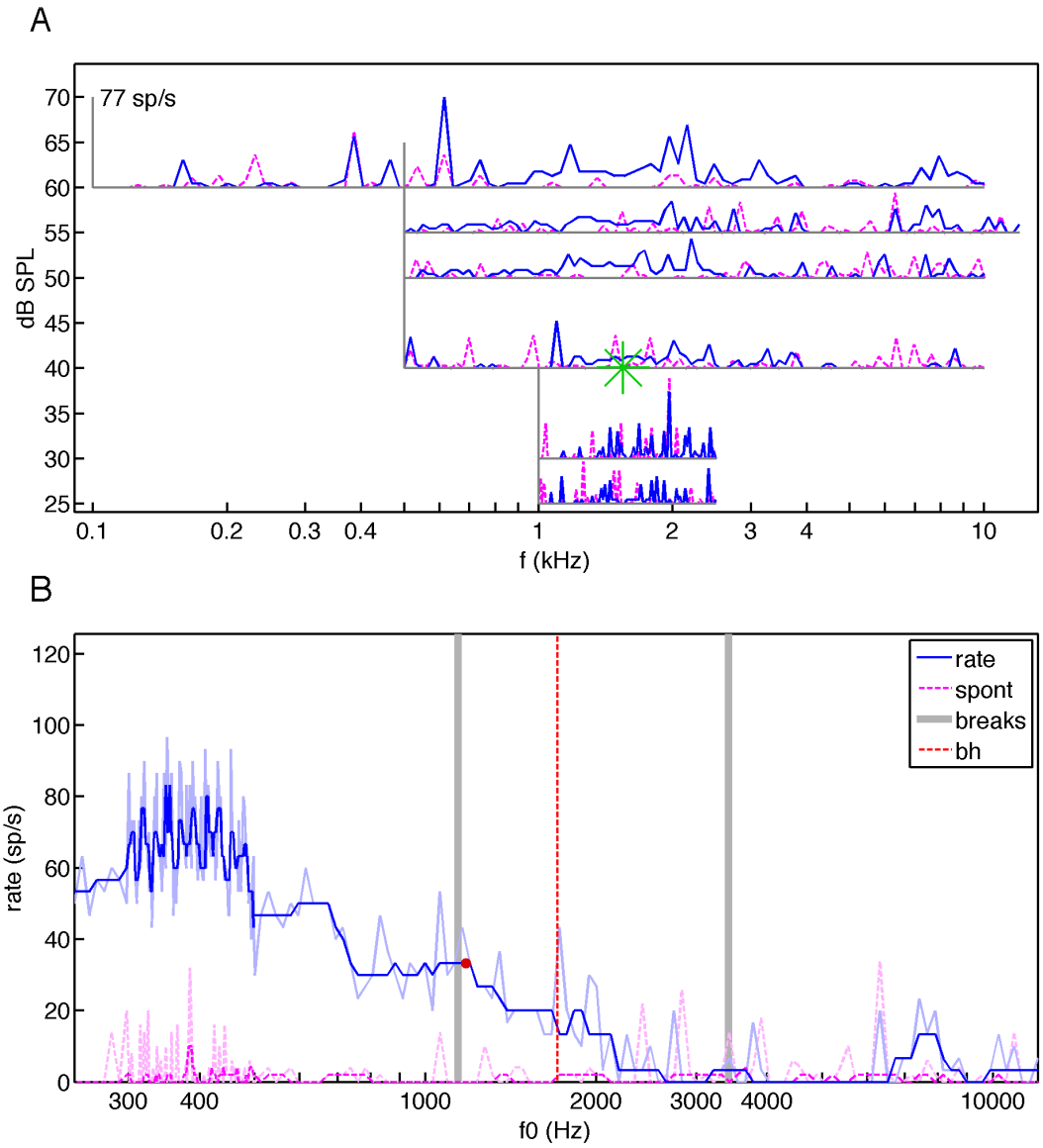

Figure 29: (o symbol in Figure 24): Like Figure 26 but for the unit with the maximum facilitation index. The unit was broadly tuned. The response is much stronger to *hm* than to tones, with a plateau around f_0 = bf/5 – bf/3.5. The only *tsh* taken for this unit was at 4.5'th order (this unit does not show up in Figure 25 since it had no integer *tsh* data). This *tsh* only showed a 15% variation between it's peak and valley, but the P_I was larger (0.26) because there was lots of noise in the *tsh*, making the minimum appear lower and maximum higher despite the median filtering.

The trend of favoring the *pure* over *hm* maps ($F_I < 0$) became more clear when we looked only at the highest order peak in the *hm* for each unit (where non-linear effects like receptive field crowding were strongest). We defined a suppression index, which was *negative* when the response to the *pure* was stronger than the *hm*, similar to the one in

(Feng, 2013: p68):

$$S_u = \frac{P_{pure} - P_{hm,n}}{P_{pure} + P_{hm,n}} \quad (13)$$

Where P_{pure} is the smoothed *pure* induced rate peak and $P_{hm,n}$ is the *smoothed n*-th order induced rate peak of the *hm,* where *n* is the highest order *hm* peak that was found (see Analyzing data on a per-order basis). A unit's S_u is identical to it's F_i at the highest order for which a *hm* peak was found. Figure 30 shows the S_u of each unit. The majority of units were suppressed as shown by $S_u < 0$ (binomial test, p = 0.004). Furthermore, there was more suppression than the linear model predicted for the majority of units (binomial test, p = 0.026). Figure 31 shows one of the many units with $S_u < 0$ for which the linear model predicted an $S_u > 0$. In this unit, as any unit, the predicted firing rate goes to infinity in the limit of zero f_0 because more and more components overlap but the actual rate saturates at a limit. In this and many other units, saturation becomes important before the order grows so high that statistical significance of the *hm* peaks is lost.

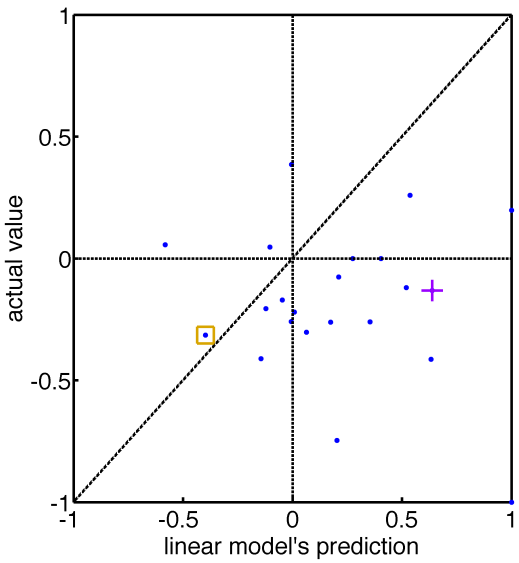

Figure 30: The actual suppression index (S_u) compared to that predicted by a linear model of all units. Values were clamped from -1 to 1.
The □ symbol corresponds the unit in Figure 22 E. This unit was the best example of a unit that had a clean response map and less suppression than predicted by the linear model. The unit had a slightly stronger response to the *hm* maps at 2-4th order than to *pure* maps (although there was inhibition at higher orders); furthermore the unit was sharply tuned enough so that the contributions didn't overlap until high order. However, units that were suppressed less than expected were the exception. The unit corresponding to the + symbol is more typical, and is shown in Figure 31.

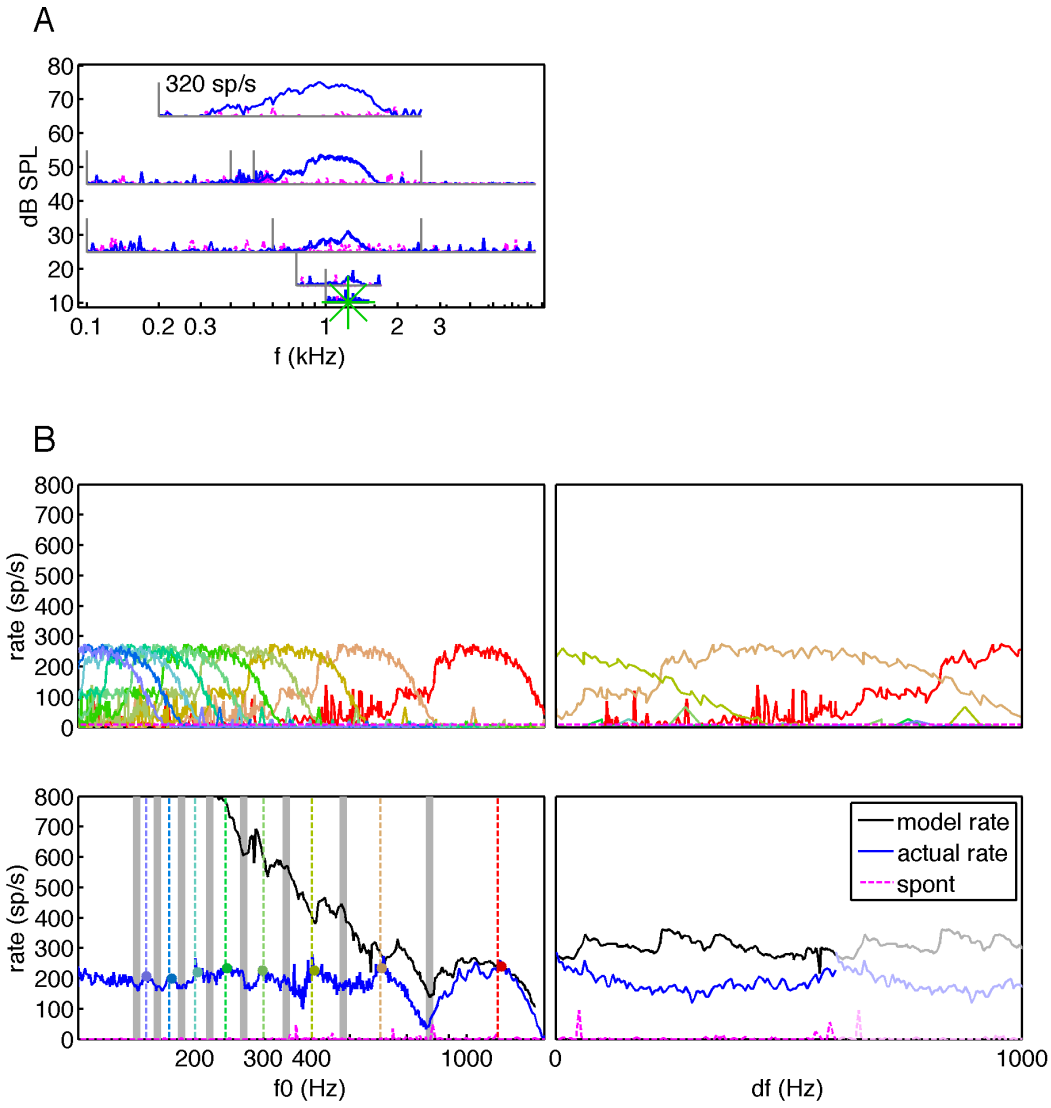

Figure 31: (corresponds to + symbol in Figure 30): A unit with more excess suppression (lower suppression index S_u) than what was predicted by the linear model.
A, The response map.
B, The linear model (displayed in the same manner as Figure 19). The linear model predicted a strong response, even at moderate orders, because the contributions were so wide that they easily overlapped. This is the case for both the *hm* and this 2nd order *tsh*. However, the actual response stayed slightly lower than the response to tones (a suppression index of -0.13) instead of growing large when there were overlapping contributions.

Tuning strength was another response characteristic that was investigated. A high facilitation index is not the only way to be sensitive to harmonicity. A strongly enhanced tuning sensitivity to the amount of mistuning (*tsh* map d_f) and/or the precise value of f_0 in

a *hm* map would be indicative of harmonic processing as well. However, no anomalous effects of tuning strength were seen. We compared the actual value of tuning vector strength for the induced rate to that the linear model. No single-order *tsh* or *hm* dataset had a significantly different tuning strength (by the binomial test). Overall the *hm* tuning strength was stronger than expected (binomial, $p = 0.01$), but this effect was small. Figure 32 and Figure 33 summarize the tuning strength data for the *hm* and *tsh*, respectively. Despite the overall trends, there were a few units that were substantially more strongly tuned than what the linear model predicted, but no "anomalous" sensitivity to harmonicity was necessary to explain this behavior. An example unit with a clean receptive field and a stronger-than-predicted tuning strength is shown in Figure 34. This unit had suppression to harmonics at 2nd and higher orders and (more importantly) maintained essentially a zero valley floor between peaks. The linear model did not capture this suppression. Instead, the components began to overlap by 2nd order, "filling in" the valleys between the orders (as what happened to the 1.5th order valley in Figure 31) and therefore predicting a low tuning strength. Suppression was the most likely cause of these departures from the linear model's predictions at low and moderate orders, while saturation was more important at higher orders.

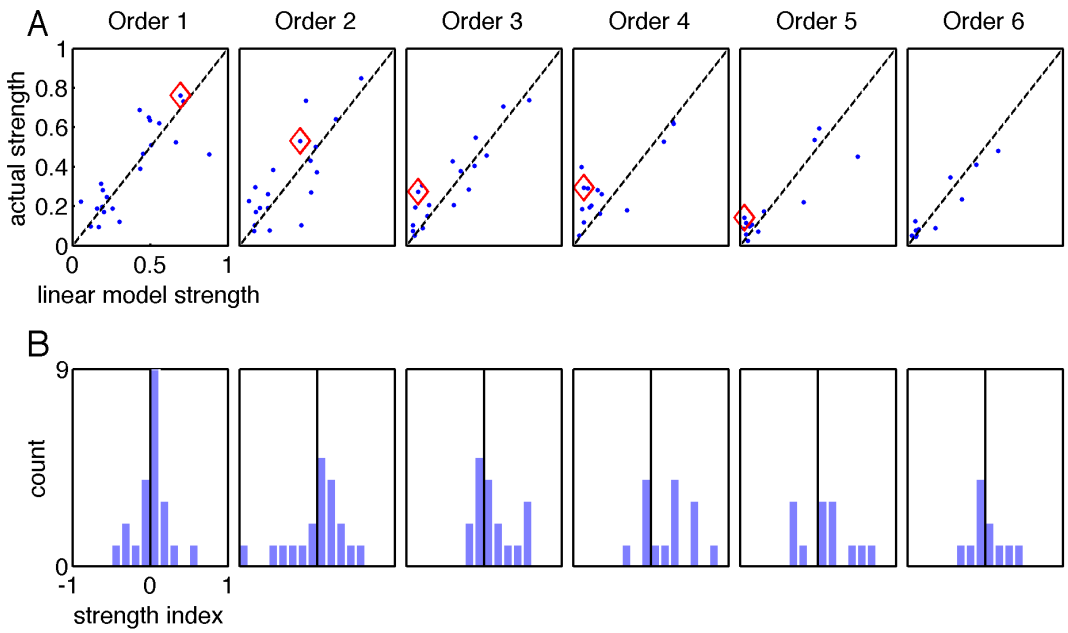

Figure 32: Actual vs linear model tuning vector strengths for *hm*s of each order up to 6th order (different subplots are different orders). The ◇ symbol visible in orders 1-5 corresponds to the unit in Figure 34.
A, The tuning strength values themselves.
B, The "strength index", which is defined identically to eq. 10 except that it used the actual vs predicted strength in place of the peak to the *hm* vs *pure*, respectively. Like the F_1, the strength index ranged from -1 (actual strength much stronger than the linear model's prediction) to 1 (actual strength much weaker).

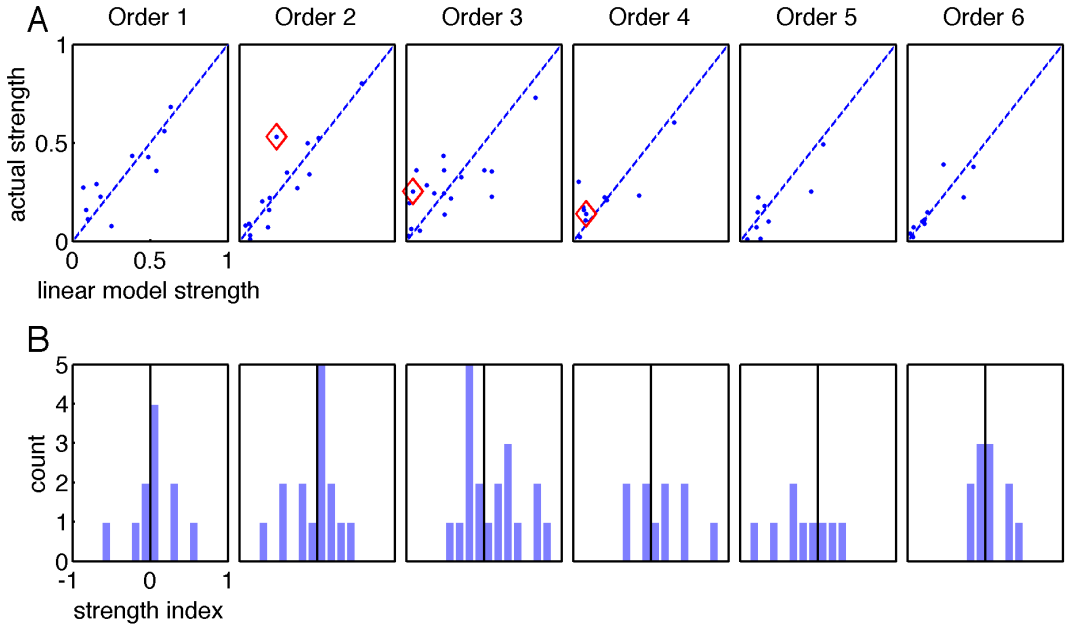

Figure 33: Similar to Figure 32 except that we are looking at the *tsh* data.

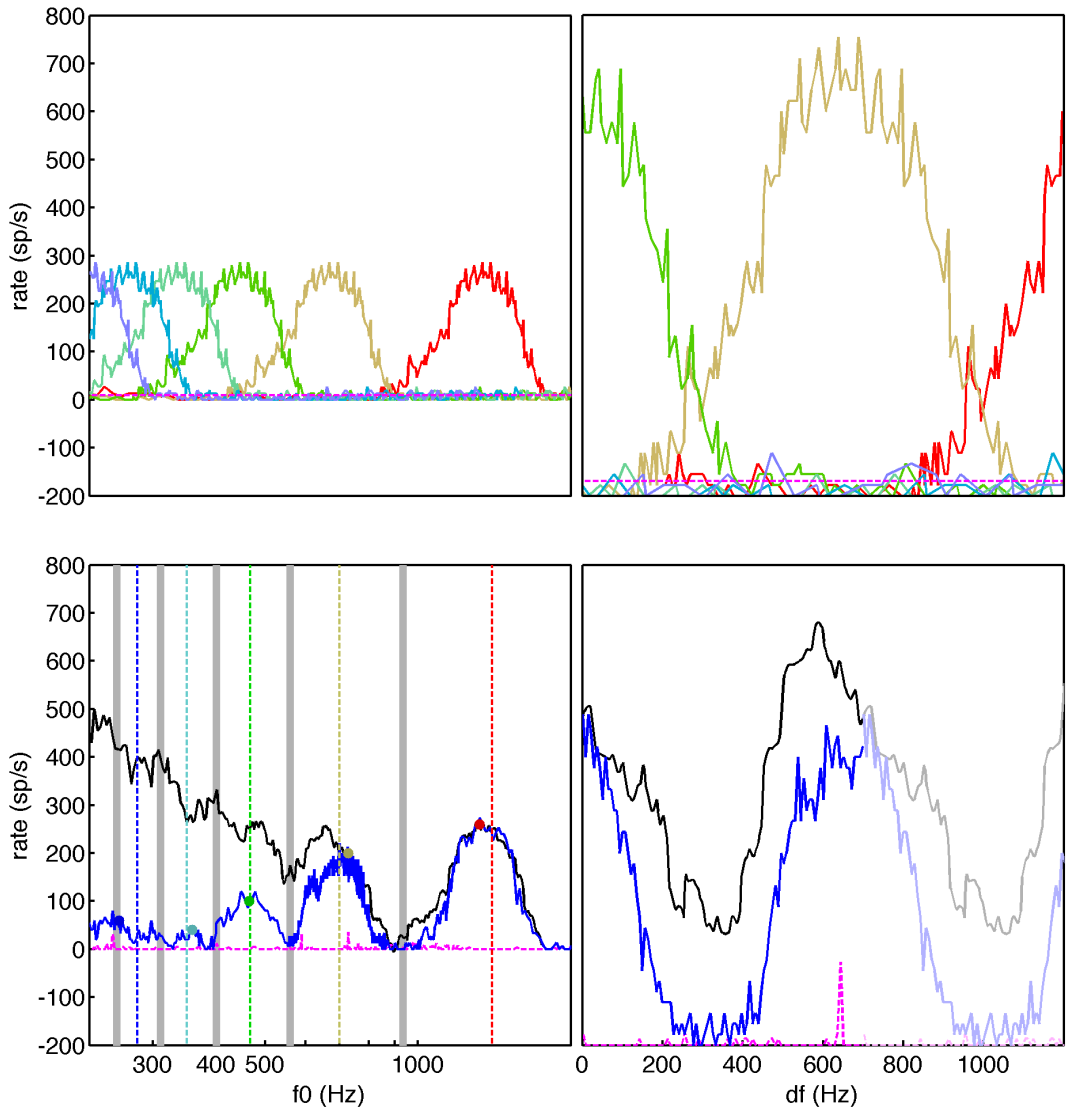

Figure 34: (corresponds to ◇ symbol in Figure 32 and Figure 33). An example unit with a stronger-than-expected tuning strength, shown in the same manor as Figure 31. *This is the same unit as in Figure 27.* In the *hm* map at higher orders there is a strong suppression in the actual (blue line) response, which is absent from the linear model's prediction (black line). For this 2nd order *tsh* map, the linear model predicts almost a uniformly higher rate than the actual response.

Harmonic template constructs

Despite the lack of harmonic selectivity of individual units, some of the template units, constructed using the pseudopopulation method in May et al. (1998) and Cai et al.

(2009), were selective in terms of the *input* current to the unit.

Of the 20 attempts to construct a template unit, 6 qualified as having a "harmonically selective" input based on the facilitation index (F_1) and periodicity index (P_1) cutoffs in (Feng, 2013: p71). Unlike the prior calculations for which we based P_1 on the maximum and minimum rates of the *tsh* (criteria which strengthened our previous null result because they overestimated the true value), for the template constructs we based P_1 on the smoothed induced rate at $d_f = 0$ vs at $d_f = f_0/2$, which are the d_fs used in Feng (2013). To further match the criteria in Feng (2013), we calculated F_1 using only the *tsh* that corresponded to mistuning a harmonic complex with f_0 equal to the engineered b_{f0} (which was very close to the actual b_{f0} for all 6 units). Figure 35 summarizes this selectivity data and how it correlates to tuning. Selective units tended to be at higher frequencies (ranksum, p = 0.0007) and higher Q_{40} values (ranksum, p = 0.0015) than non-selective ones. A unit for which it's template construct had an input well within the range of being a "harmonic template: Figure 36. Figure 37 shows an almost successful template construct for a unit with a moderate b_f. Both of these examples had sharply tuned responses to *pure*, *hm* and *tsh* stimuli that allowed the template construct to be sensitive to the location of each component in a given stimulus.

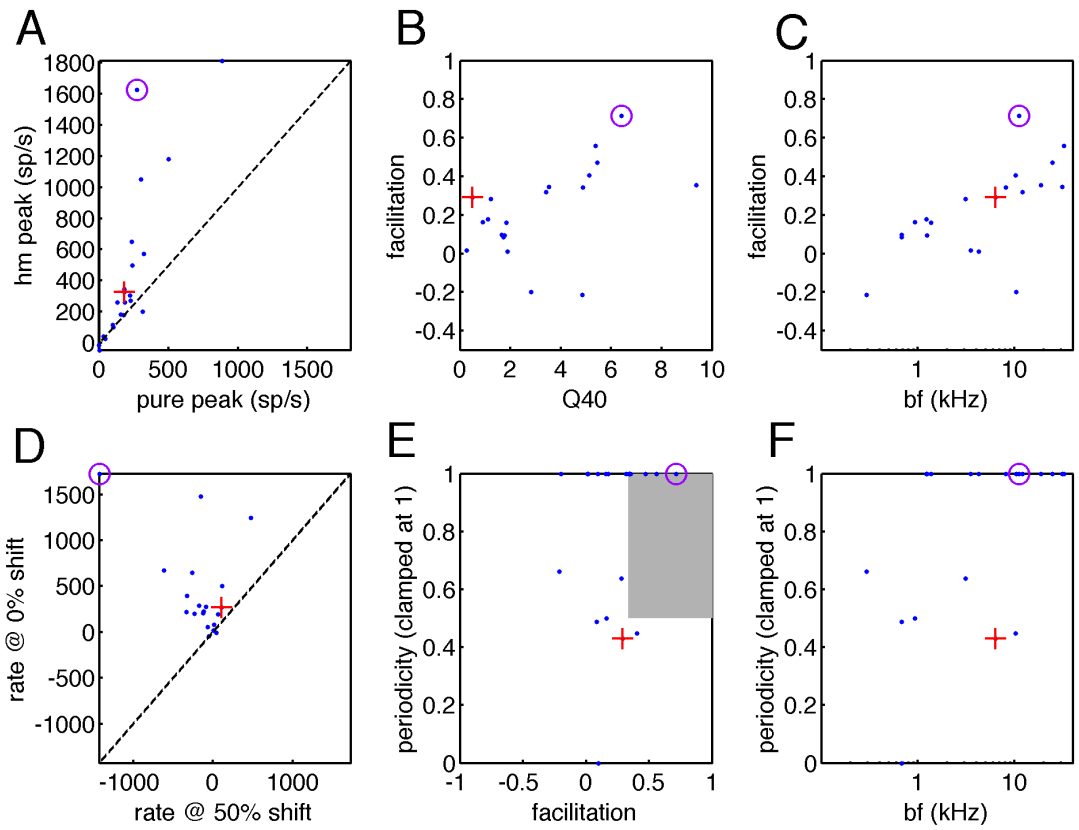

Figure 35: Results of harmonic template constructs. Colored symbols correspond to Figure 36 and Figure 37. Two bad data points (insufficient data) which were in the selective region were removed. Some units with marginal data quality remain in the "non-selective" pool, so the actual proportion of template units may have been *higher* had we used a more stringent data quality cutoff.

A, The peak of the *pure* vs *hm* for each template unit.

B, Q_{40} for the real units vs facilitation index (F_1) for the template units.

C, b_f for the real units vs F_1 for the template units.

D, Peak input for harmonic vs fully-mistuned *tsh* tokens for the template units.

E, F_1 vs periodicity index (P_1). The shaded rectangle in the upper right corresponds to the region that qualified units as "template units" in Feng (2013: p71). The P_1 was clamped between 0 and 1.

F, b_f for the real units vs P_1, again clamped between 0 and 1.

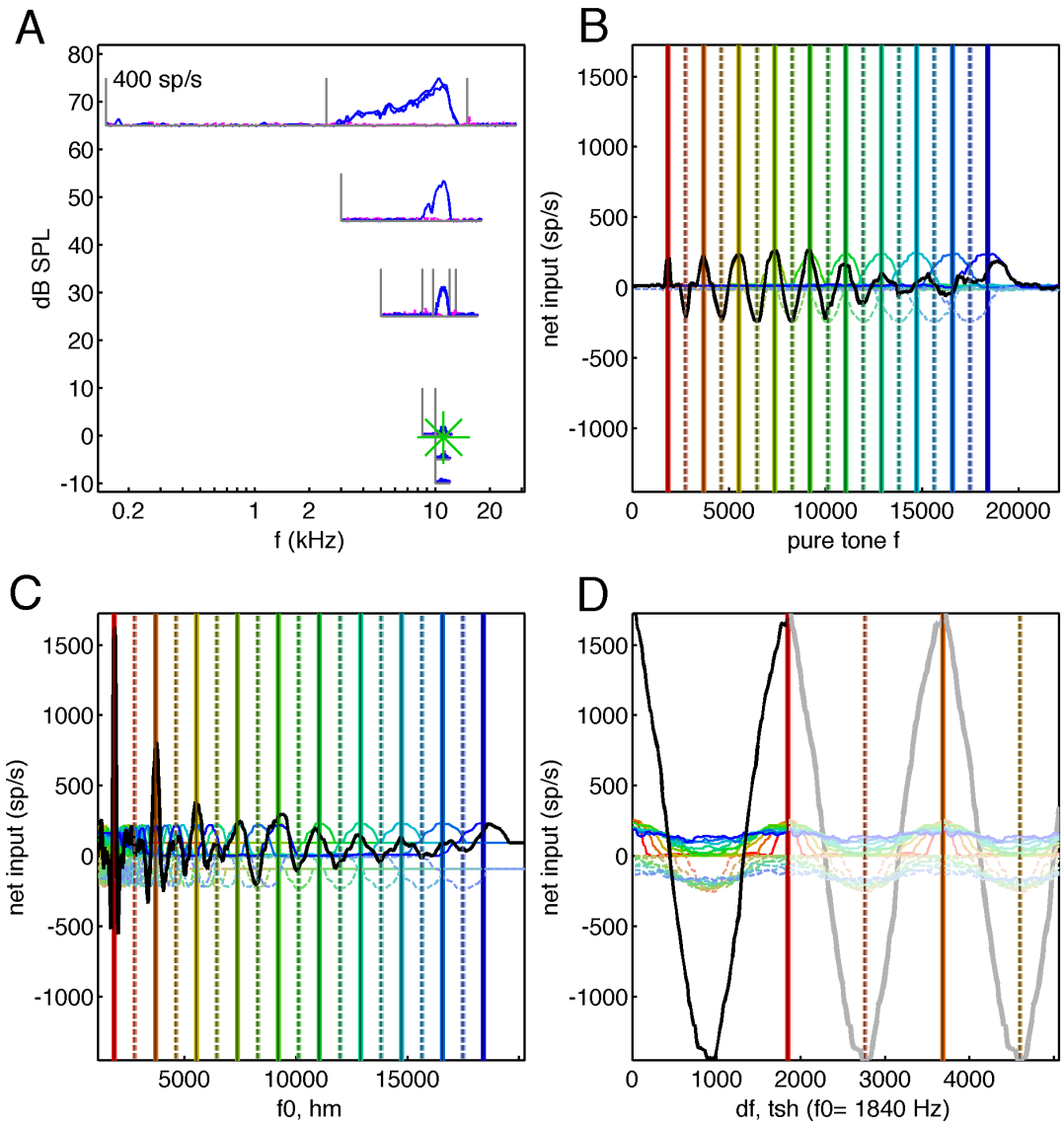

Figure 36: (corresponds to the o symbol in Figure 35): A constructed template unit which had a harmonically selective net input ($F_I = 0.71$, $P_I > 1$), it was constructed from a unit with a high b_f of 11.0 kHz.

A**,** The response map for the unit as in Figure 14.

B,C, and D: Figure 20C-style template input to a *pure*, *hm*, and *tsh* maps, respectively. Like in Figure 20C, each component is shown as a colored curve. The total current into the template is shown as a black curve. The pseudo-b_fs are indicated by vertical bars: solid is excitatory and dashed is inhibitory. The b_{f0},1.84 kHz (red vertical bar all the way to the left), is equal to the spacing between adjacent excitatory (or adjacent inhibitory) bars. The *hm* had a very strong, sharp response at b_{f0}. The *tsh* in D had $f_0 = b_{f0}$, like any *tsh* for a template construct. The *tsh* shows one cycle in strong colors and replicates the cycle in faint colors.

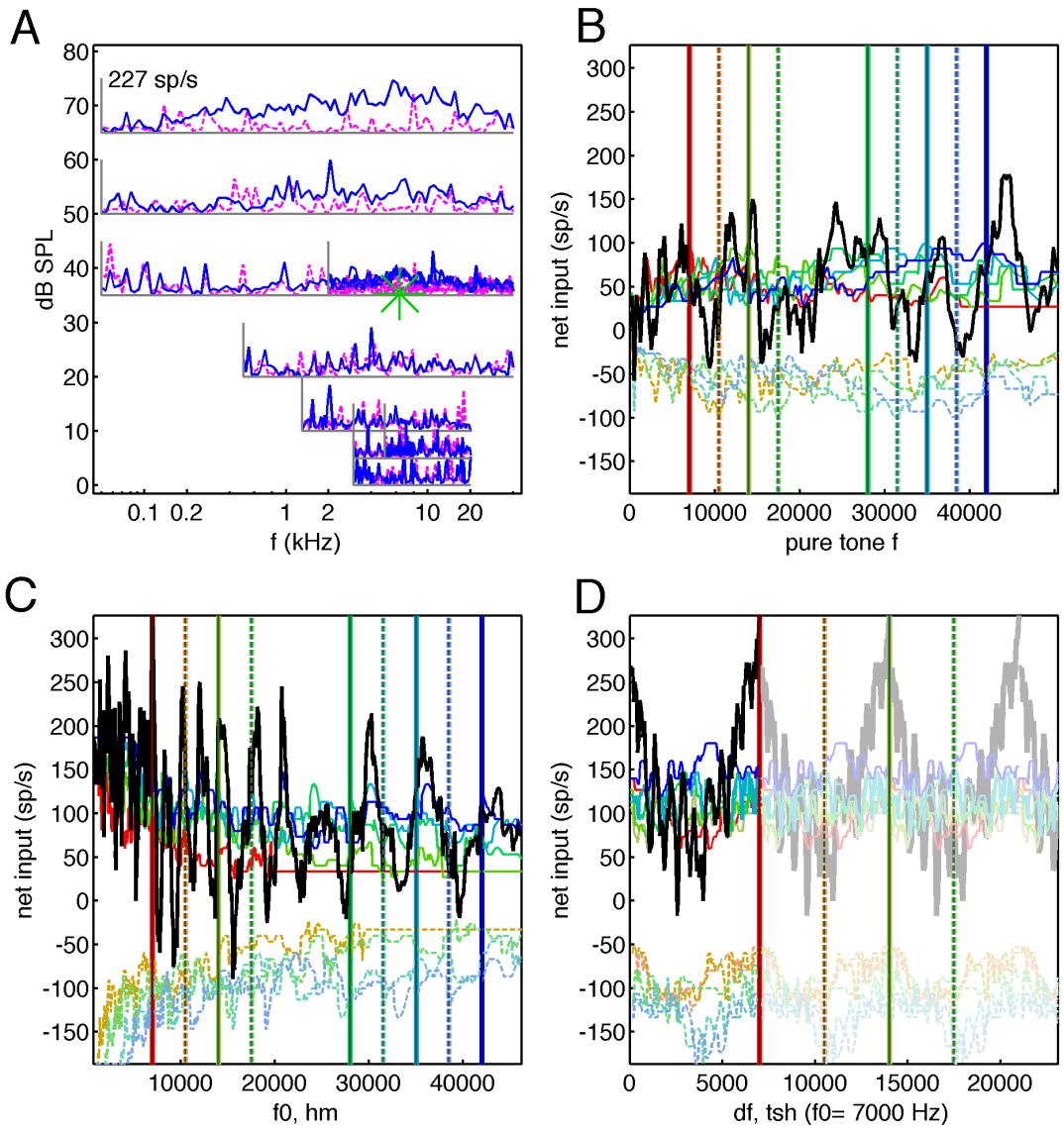

Figure 37: (+ symbol in Figure 35): Like Figure 36 but for a unit with a moderate best frequency (7kHz) and almost qualifying for harmonic selectivity ($F_I = 0.29$, $P_I = 0.43$) in terms of input. The patterns of input to the three maps weren't nearly as clean. However, the *hm* map in C still had a sharp global maximum at b_{f0} and the *tsh* in D still showed sensitivity to mistuning. There are "missing" pseudo-b_fs because isolation was lost before all the *tsh* data could be collected.

By excluding the lower frequency pseudounits we can change the shape of the template's input current. Instead of having a single, strong peak at b_{f0} we can make the current have several peaks that are similar heights around b_{f0}. Figure 38 shows the result of restricting the inputs to 3 excitatory and 2 inhibitory ones, using the same unit as was

used for Figure 36. The construct's input was still “harmonically selective” as per Feng (2013) but input on the *hm* maps had three nearly equal peaks.

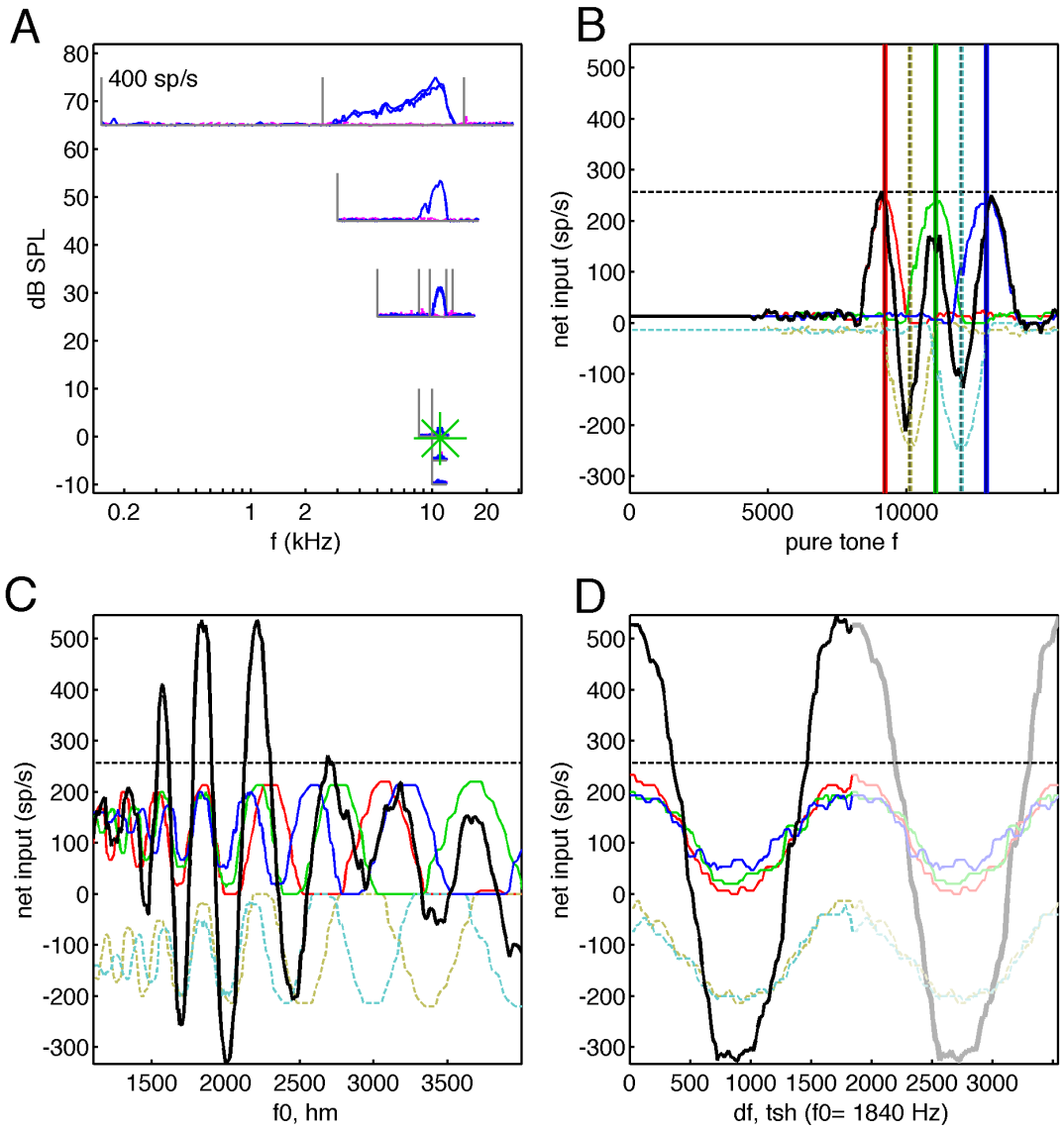

Figure 38: The template constructed using the *same unit* as Figure 36, but restricting the inputs to excitation at orders 5-7 and inhibition at orders 5.5 and 6.5. This combination of orders produced three peaks in the *hm* that are similar height and still are considerably stronger than the peak of the *pure* map (black dotted line). The input's F_1 was 0.35 and P_1 was above 1, which fit the criteria in Feng (2013) for being harmonically selective.

The same process of constructing a template unit from the Marmoset units was used for the model cat auditory nerve (AN) units. We picked the same reference orders as

the example in Figure 36: excitation at orders 1-10 and inhibition at orders 1.5-9.5. Stimuli were presented 40dB above threshold. Due to the tightly controlled parameters, the "measurements" from different units were much more consistent. Apart from a reduced noise, the findings were similar to the marmoset units: higher b_f units had a higher Q_{40} value and a sharper peak pattern in the *hm* maps. Also, the input to the template constructs had higher facilitation index and periodicity index at higher b_f. These results are summarized in Figure 39. Examples of units with different b_fs are shown in Figures 40-42. These examples showed a similar trend as the marmoset units: the higher frequency units made more harmonically selective template constructs.

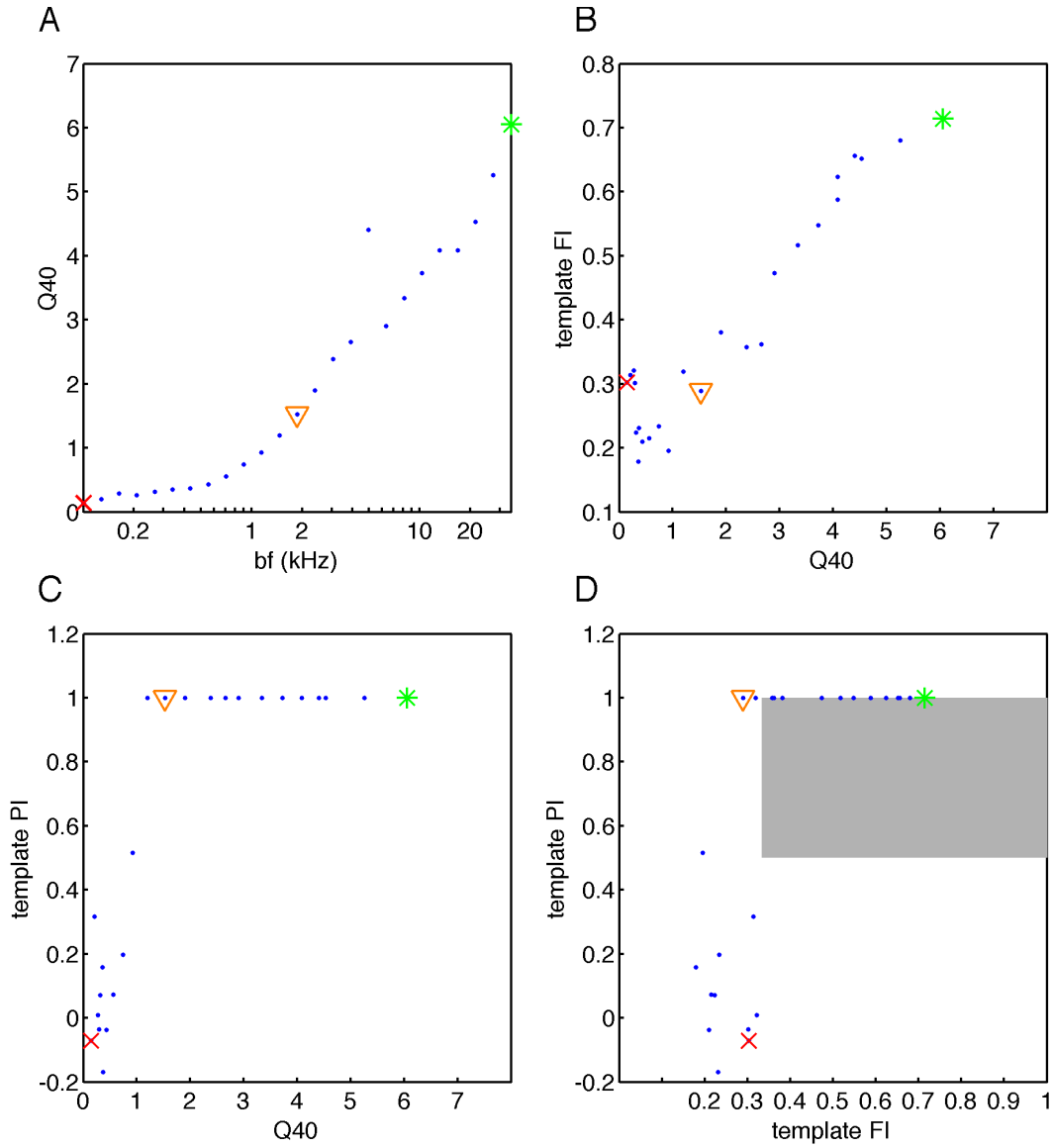

Figure 39: The cat auditory nerve model. Highlighted symbols show examples in Figure 40-Figure 42.
A, the Q_{40} vs the b_f for the simulated units shows a clear correlation (the outlier with an unusually high Q_{40} was near the spectral notch in Figure 21).
B, Q_{40} vs the facilitation index of *template units* constructed from on the simulated auditory nerve units. There is a strong linear correlation (adj r^2 = 0.90).
C, Q_{40} vs periodicity index of the input to the template units, P_1 is clamped below 1. Above a Q_{40} of about 1.5 (around 2 kHz), the P_1 jumps to >1.
D, The F_1 and P_1, for inputs to the template units. Units in the shaded region have an input that is considered harmonically selective according to Feng (2013): p71. The cut-off b_f, above which the inputs were harmonic, was about 2.5kHz.

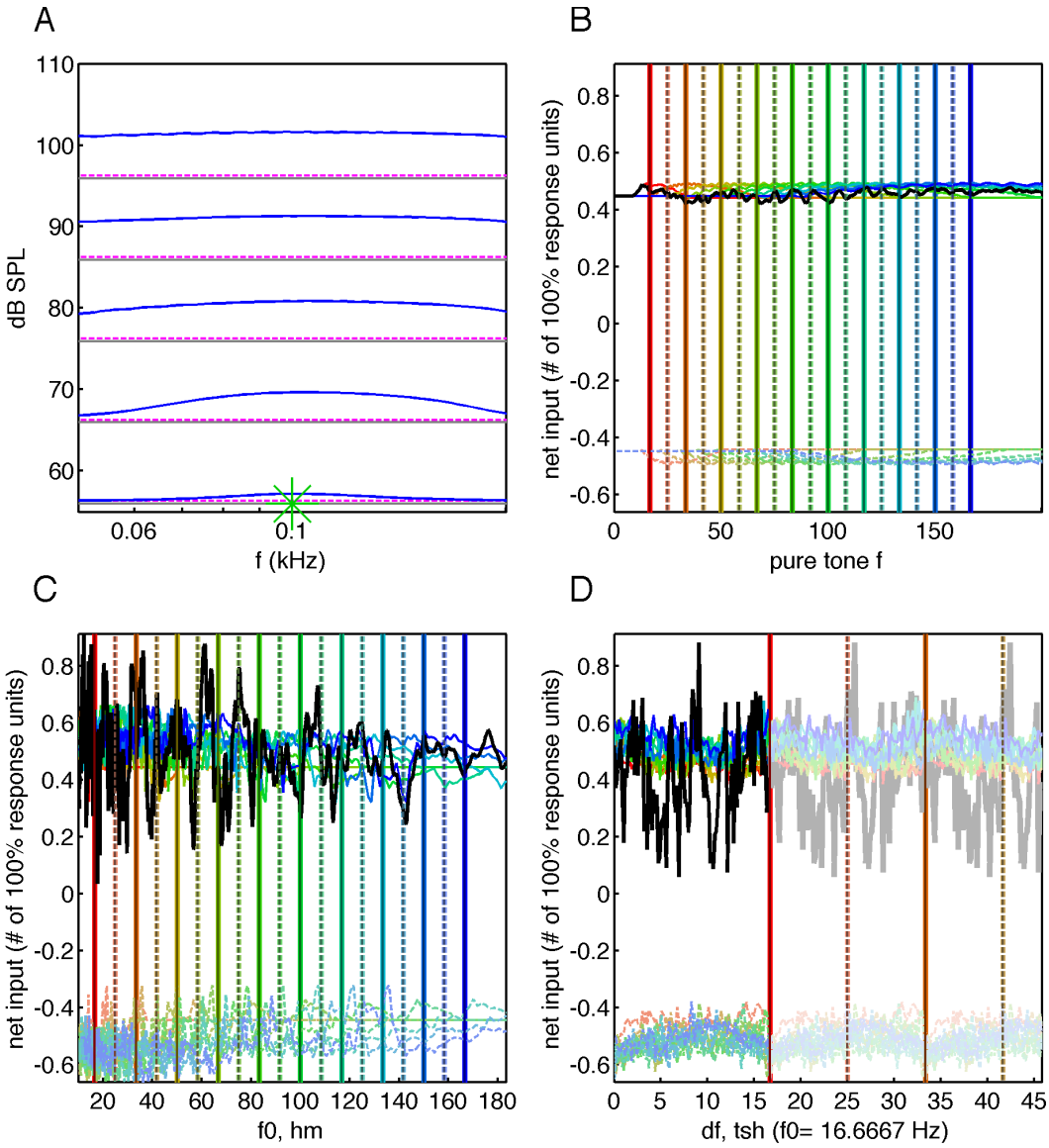

Figure 40: (corresponds to x symbol in Figure 39): The same as Figure 36 but for a *low* frequency (b_f = 0.1kHz) model auditory nerve unit. The input was almost constant throughout the *pure hm* and *tsh* maps, however there was still a little bit of selectivity, as indicated by the slightly stronger response to $f_0 = b_{f0}$ in the *hm* and the $d_f = 0$ in the *tsh* map.

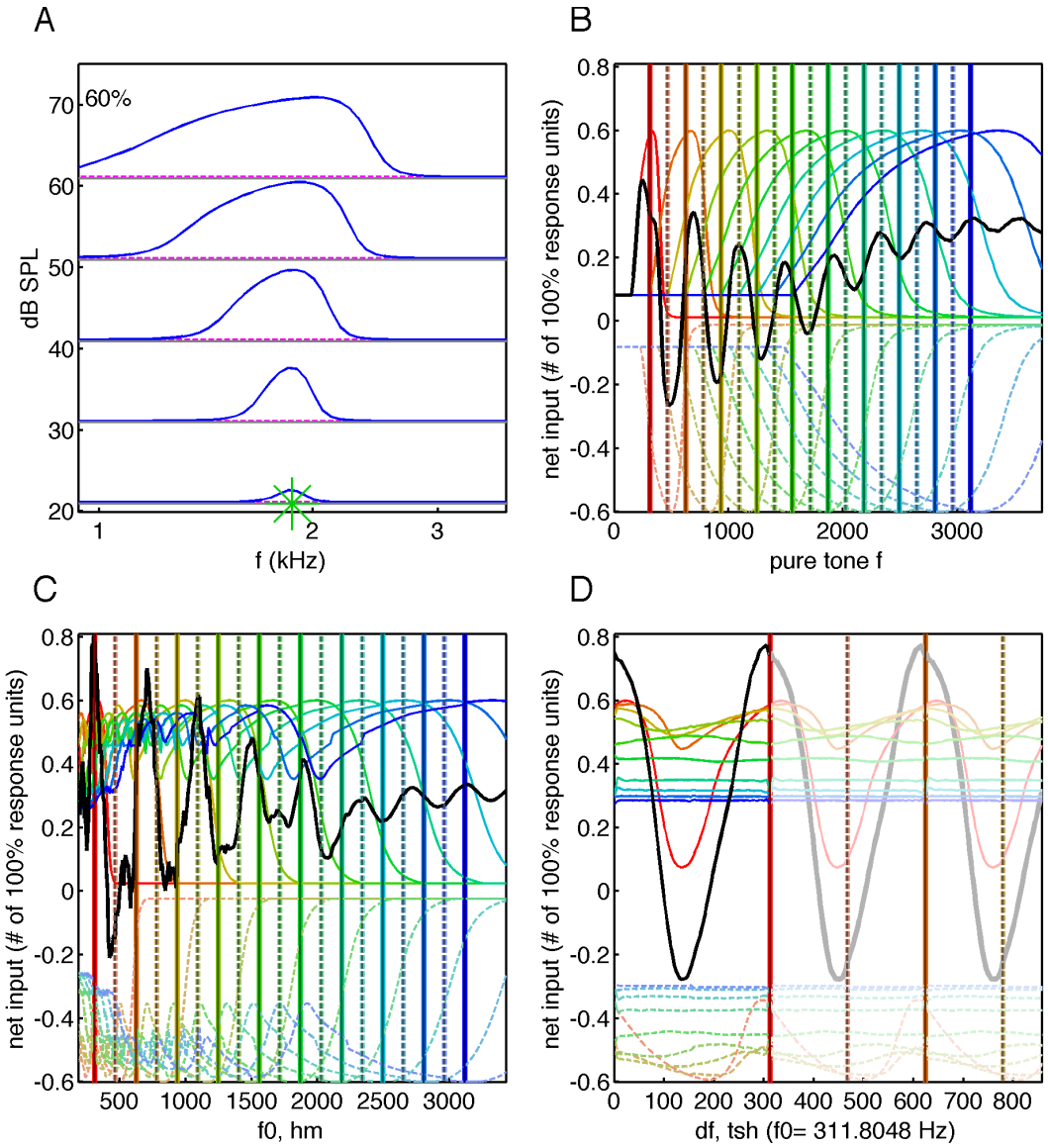

Figure 41: (▽ symbol in Figure 39): Same as Figure 40 but for a medium frequency (1.87kHz) unit. The sensitivity for tuning of the input to the template was almost enough to quality as harmonically selective.

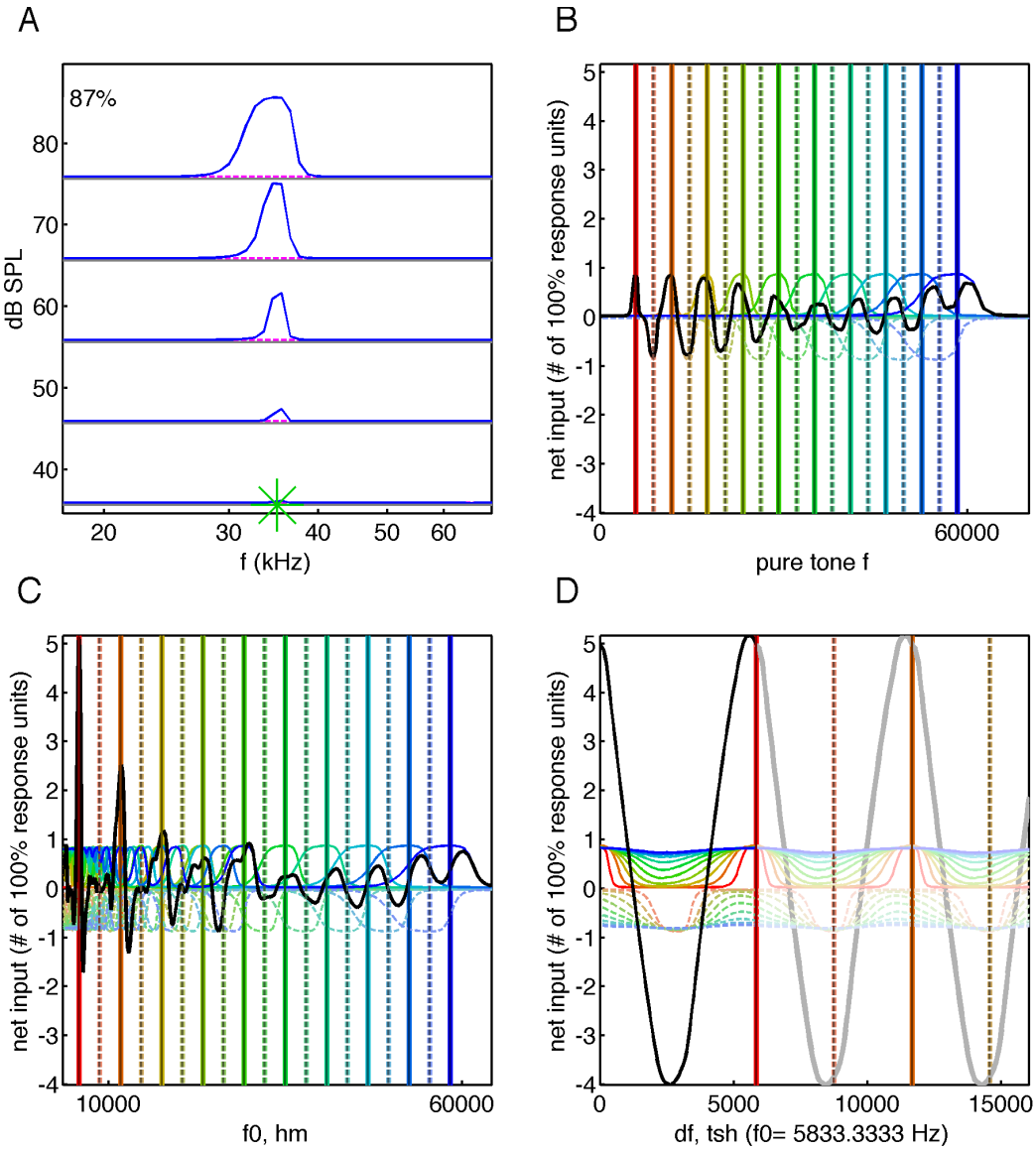

Figure 42: (* symbol in Figure 39): Figure 40 but for a high b_f (35kHz) unit. The input to the template construct was very sensitive to tuning, particularly to changes in f_0, and was well within the criteria to be considered "harmonically selective".

Discussion/Conclusion

No harmonic selectivity was observed in the ICC

The marmoset ICC probably does not have "harmonic template units" at a physiologically relevant level. The cutoff for a "template unit" is a facilitation index (F_i) of at least 1/3 along with periodicity index (P_I) of at least 0.5 (Feng, 2013: p71); 64/559 = 11% were classified as "template units" in Feng (2013: p63). None of our 22 units fit these criterion, but we can't conclusively rule out template occurring at the 11% level (binomial test, p=0.069) on this metric alone. However, Feng (2013: p48) had a continuum of F_is. Our data, *among units with sensitivity to mistuned complexes*, topped out at an average $F_I = 0.04$. Furthermore, Feng (2013: p48) had $F_I>0$ for the majority of the units but our results showed a trend toward $F_I<0$. For these reasons, in addition to an almost significant result, it seems unlikely that template units exist in the ICC at a meaningful level.

Most units showed selectivity to f_0 and mistunings for a given harmonic complex, but the linear model predicted similar tuning properties, which indicates that this effect can usually be explained as frequency tuning, i.e. to sensitivity to the presence of a component near to or at b_f. Both the ICC units here and the template units Feng (2013) (such as the one in Figure 4), were selective for f_0. Most of the ICC units responded whenever there was a component at b_f, no matter the stimuli. This behavior was in contrast to the template units in Feng (2013), which were also sensitive to components

away from b_f and responded much more weakly when the non-b_f components in a harmonic complex were jittered (Feng, 2013: p56).

Our model was very limited in the amount of inhibition it can capture: the input rate can not be more negative than the spont, and in most cases the spont was small. A side-band inhibition, as in Winer and Schreiner (2005: p315) could reduce the range of excitatory frequencies, as seen in Winer and Schreiner(2005: p314). Such an inhibition, if strong enough, would not be captured by the linear model. Furthermore, the model predicted an unrealistically high firing rate for spectrally dense stimuli, while the actual rate saturated. Both of these effects, observable in Figure 31, accounted for the units that had a much sharper response to *hm* or *tsh* than "expected" based on the linear model: by lowering the absolute firing rate, the tuning strength, which is based on the *relative* rates, was made larger. Overall, these units were still responding to the presence vs absence of a sound at b_f. Despite these non-linear effects, the overall behavior of most units was still essentially dependent on whether there was energy in the receptive field and not to harmonicity *per se*.

The mysterious world of the ICX

Our study only worked with the central nucleus of the IC (the ICC), but there are two types of units in the external nucleus of the IC (ICX) that may respond more strongly to harmonics. During most of the searching, the hash in the ICX strongly favored noise over tones, which is in agreement with observations that ICX units tend to be broadly

tuned (Pickles, 2008: p192). Preference for multi-component stimuli could indicate units that respond more strongly to *hm* maps than to *pure* maps. A generic many-input neuron would produce results similar to our high F_1 units: despite responding to harmonics more than tones, it would be poorly sensitive to mistuning or to the value of f_0 in an *hm* token (as in Figure 29). These wouldn't be considered "template" units because they would have a periodicity index near 0 since they lack the resolution to differentiate between harmonic and mistuned harmonic complexes.

A more interesting property of the ICX is that it receives the bulk of the descending input (Winer and Schreiner, 2005: p241). A direct 1:1 excitation from a template cortical unit would create a *true* harmonic template unit in the ICX. The presence of such units would suggest a physiological function for harmonic selectivity in the descending pathway. More investigation with this region, perhaps in an acute rat model for which the ICX can be directly exposed, is warranted.

Some high frequency ICC and AN neurons can be used to build harmonic template neurons

The tuning sharpness of the AN and ICC neurons was similar: the Q_{40} ranged from about 0.5-6 for typical units (see Figure 39) and increased across the hearing frequency range. The experimental ICC data was much more noisy than the idealized AN model, of course. The predominance of selective template constructs at higher frequency is probably due to the higher Q_{40} values in both the AN and ICC units. Sharp tuning was important for resolving different components of harmonic sounds, which made a more

effective spectral sieve.

Template units could be created by summing up excitatory and inhibitory input from either ICC and AN neurons at correctly positioned frequencies, provided that the feed-in pseudounits had sufficiently sharply tuning. The templates in our model were excited by components at $k_{excite}b_{f0}$ and inhibited by components at $(k_{inhibit}+0.5)b_{f0}$. A *hm* token with $f_0 = b_{f0}$ had components in the excitatory regions but no components in the inhibitory regions. However, if *either* the f_0 changed or the complex got mistuned (i.e. a *tsh* with $d_f>0$), the template unit got less total input because the components were no-longer at the correct frequencies. The *hm* peak was also higher than the *pure* peak since only one feed-in at a time could be strongly driven.

The details of the template's response will depend on the relative threshold of a unit as well as which k_{excite} and $k_{inhibit}$ the unit receives. Template units could use a high threshold to sharpen their response, provided the maximal *input* was still stronger to harmonic complexes than either to tones or mistuned complexes. This would allow the output of a template unit to be "harmonically selective" even if the input wasn't. Another effect that adjusts the shape of the response is the range of *k* values. Units with a wide range that require most/all of their excitatory inputs to be energized in order to fire will tend to have a well defined, narrow response to harmonics with a single strong peak at b_{f0} (similar to the inputs in Figure 36 but with a threshold below which no spikes occur at all). Units with a narrower range of *k*-values may look more like the multi-peak curve in Figure 38 which has three nearly equal peaks in the *hm* response. The mechanism behind these two effects is explained in Figure 43.

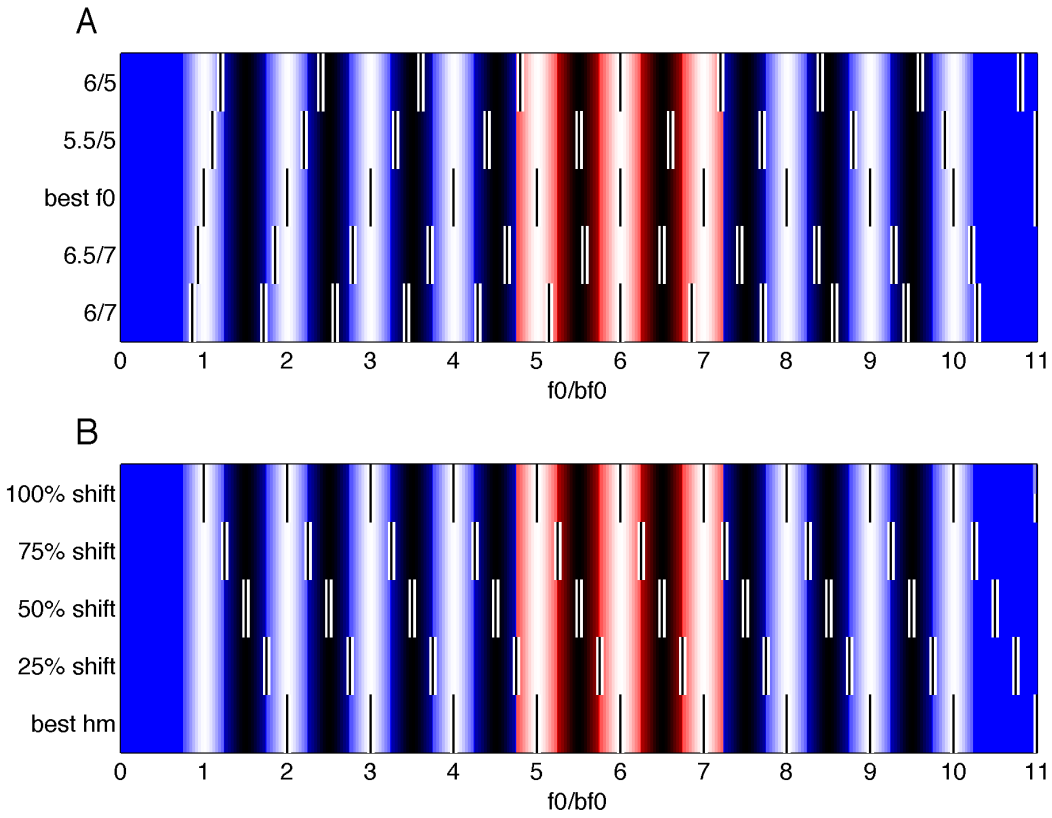

Figure 43: Illustration of how a template unit would see different stimuli. Each row of thin bars represents the components of a token.
A, different *hm* tokens.
B, different *tsh* tokens.
The background represents the receptive field of a "wide" and "narrow" template unit. The lighter areas are excitation and the darker areas are inhibition. The entire area, including both blue and red regions, is the range of inputs used in the "wide" template unit in Figure 36. The red area alone represents the "narrow" template used in Figure 38 that only had three excitatory inputs. Suppose both template units need to have all the excitatory regions and none of the inhibitory regions activated in order to fire. Both would be *bona fide* template units since they will not fire to tones (since tones can only energize a single area) or when the harmonic complex is mistuned (in our example, mistuned by 25% or more) because the components will be shifted into the inhibitory regions instead of the excitatory regions. The wider template will only respond to harmonics when the fundamental frequency f_0 is very close to b_{f0}; even a slight shift will cause the components on the edges of the sieve to miss the excitatory areas. The narrower template, on the other hand, will fire to harmonics when f_0 is near $6/7b_{f0}$, b_{f0} or $6/5b_{f0}$, but not in between. This will give it have 3 peaks in the *hm* map. The narrow unit is more flexible to spacing than the wide unit but is still a template.

Difficulties with temporal sensitivity due to tuning

Time-domain harmonic sensitivity, such as that needed for the pitch of unresolved harmonics, was not investigated in this study. The data acquisition used harmonic stimuli,

which were usually resolved in the frequency domain, not click trains (which would be resolved in the time-domain). Time-domain-sensitive units would probably tend to be found at lower frequency because low b_f AN units have poorer frequency resolution (lower Q_{40} values, see Figure 39A) and there is a bias for toward low-frequencies in pitch-sensitive cortical units (Bendor and Wang, 2010). Unfortunately, there was a lack of low frequency but high Q_{40} ICC units in this study (the one exception had very strong inhibition and noisy, below-spont responses to *hm* maps). Instead, the pattern of higher Q_{40} in higher frequency IC units followed that of the cat AN (Pickles, 2008: p80), indicating that most of our units were "AN-like". This did not mean that the animal lacked low b_f, high Q_{40} units in the ICC. Our sampling technique, like the majority of non-cortical primate electrophysological studies, was designed to get data rather than systematically explore every region of the IC central nucleus (ICC). The non-uniform samples, "tracks" through the ICC, may have missed low-b_f-high-Q_{40} regions. The IC is not homogenous in the two directions orthogonal to tuning. Within each frequency lamina there are gradients in the number of projections from the dorsal vs ventral cochlear nucleus and other feed-in structures such as the lateral superior olive (Pickles, 2008: p81). Other studies showed clustering of inputs, indicating that there may be microcircuits specialized for processing specific information (Pickles, 2008: p82). Finally, the response of units in the dorsal region may be "more complex" as indicated by latency and anatomical findings (Pickles, 2008: p82). The significance of these differences within a lamina, while probably important, is poorly understood (Pickles, 2008: p82). Since selective, low frequency units were would have been needed for

investigating time-domain effects, no such effects were found.

Constructing a time-domain model, which would be useful for low frequencies, is not as straightforward because there are many ways to produce sensitivity. It would probably be based on delay lines, such as those in Cheveigne (1998). The output of a delayed-coincidence detector would pass through some sort of thresholding circuit in order to prevent “false alarms” from triggering the system. Constructing a model that takes inputs from simulated AN fibers would be possible, but with so many free parameters it would hardly be meaningful without experimental data to support it.

High-order non-trivialities

The linear additive model with which we compared the responses with to look for any enhanced selectivity was not an accurate model of a typical neuron. The model had no saturation and did not allow strong inhibition (since firing rates can't go below zero). Saturation could be included without adding any degrees of freedom by limiting the firing rate to the unit's maximum firing rate. However, this relies on the assumption that the stimulus to near-optimally drive the unit has been found already. Two tone tests, where one tone is kept at b_f as described in Feng (2013: p29) would be useful in quantifying how strong inhibition is. Inhibition itself is more of a gate than a simple subtraction from a firing rate: a “small” inhibition can hide what would otherwise be a large excitation (Winer and Schreiner, 2005: p299). If we included saturation, gating inhibition, and other “basic” non-linearities in the model a more subtle non-linearity may

have persisted that would be useful for harmonic template processing. However, this seems unlikely. Apart from suppression effects, we failed to find any enhancement of the tuning sensitivity to mistuning or changing the f_0 in harmonic complexes. A more complex model would, if anything, *remove* discrepancies between the tuning strengths, not bring out some latent non-linearity.

Multiple sounds

All harmonic stimuli corresponded to a sound with a single f_0. However, the auditory system excels in sorting harmonic sounds with different f_0s into individual auditory objects (as discussed in Bregman, 1990). It is feasible that the neuron may have a "latent" selectivity for a particular f_0: the unit would respond to energy in the receptive field over a broad range of f_0s but would have a much narrower range of preferred f_0s when presented a "two f_0" test. If the sharpening of the "best f_0" was similar to the two-tone sharpening in Winer and Schreiner (2005: p314) and was much stronger than the suppression-based sharpening we observed, this could be useful for sound separation. More investigation into multi-sound stimuli is needed.

References

American Standards Association (1951) Acoustical Terminology. New York.

Attias H, Schreiner CE, Keck WM (1998) Coding of naturalistic stimuli by auditory midbrain neurons. Advances in Neural Information Processing 10:103-109.

Bendor D, Wang X (2010) Neural Coding of Periodicity in Marmoset Auditory Cortex. J Neurophysiol 103(4):1809-1822.

Bregman AS (1990) Auditory scene analysis: the perceptual organization of sound. Cambridge, MA: MIT Press.

Cai S, Ma WL, Young ED (2008) Encoding Intensity in Ventral Cochlear Nucleus Following Acoustic Trauma: Implications for Loudness Recruitment. Journal of the Association for Research in Otolaryngology 10:5-22.

Cariani PA, Delgutte B (1996) Neural Correlates of the Pitch of Complex Tones. I. Pitch and Pitch Salience . J Neurophysiol 76(3):1698-1716.

Cedolin L, Delgutte B (2010) Spatiotemporal representation of the pitch of harmonic complex tones in the auditory nerve. J Neurosci 30(38):12712–12724.

Cheveigne A (1998) Cancellation model of pitch perception. J Acoust Soc Am 103(3):1261-1271.

Darwin CJ, Carlyon RP (1995) Auditory grouping. In: Moore BCJ (ed) Hearing. London: Academic Press.

Feng L (2013) Spectral Integration And Neural Representation of Harmonic Complex Tones in Primate Auditory Cortex. Johns Hopkins University.

Fletcher N, Rossing T (2010) The Physics of Musical Instruments, 2nd ed. New York, NW: Springer-Verling.

Heinz GM, Swaminathan J (2009) Quantifying Envelope and Fine-Structure Coding in Auditory Nerve Responses to Chimaeric Speech. Journal of the Association for Research in Otolaryngology 10(3):407-423.

Hemmen JL (2013) Vector strength after Goldberg, Brown, and von Mises: biological and mathematical perspectives. Biol Cybern 107(4):385-396.

Kandel E, Schwartz J, Jessel T, Siegelbaum S, Hudspeth A (2012) Principles of Neural Science, Fifth Edition. New York, NY: McGraw-Hill.

Kaw A, Barker C, Integration. Holistic Numerical Methods Institute, http://www.mpia-hd.mpg.de/~mordasini/UKNUM/integration.pdf

Licklider JCR (1954) "Periodicity" Pitch and "Place" Pitch. J Acoust Soc Am 26:945.

Lu T, Liang L, Wang X (2001) Temporal and rate representations of time-varying signals in the auditory cortex of awake primates. Nature Neuroscience 4:1131–1138.

Lyon RF (1990) Automatic Gain Control In Cochlear Mechanics. Springer-Verlag 87:395-402.

Malmierca MS, Merchán MA, Henkel CK, Oliver DL (2002) Direct projections from cochlear nuclear complex to auditory thalamus in the rat. J Neurosci 22(24):10891-10897.

May BJ, Prell GS, Sachs MB (1998) Vowel Representations in the Ventral Cochlear Nucleus of the Cat: Effects of Level, Background Noise, and Behavioral State. J Neurophysiol 79(4):1755-1767.

Nelson PC, Smith ZM, Young ED (2009) Wide dynamic range forward suppression in marmoset inferior colliculus neurons is generated centrally and accounts for perceptual masking. J Neurosci 29(8):2553-62.

Osmanski MS, Wang X (2011) Measurement of absolute auditory thresholds in the common marmoset (Callithrix jacchus). Hear Res 277(1-2):127-133.

Pickles JO (2008) An Introduction to the Physiology of Hearing, Third Edition. Waltham, MA: Academic Press.

Ramachandran R, Davis KA, May BJ (1999) Single-Unit Responses in the Inferior Colliculus of Decerebrate Cats I. Classification Based on Frequency Response Maps. J Neurophysiol 82(1):152-163.

Scheffers, Maria MT (1983) Sifting vowels: Auditory pitch analysis and sound segregation. Rijksuniversiteit te Groningen. Full text: http://dissertations.ub.rug.nl/faculties/science/1983/m.t.m.scheffers/?pLanguage=en&pFullItemRecord=ON

Schroeder M (1970) Synthesis of Low-Peak-Factor Signals and Binary Sequences With Low Autocoorelation. IEEE Transactions on Information Theory 16:85-89.

Slee S, Young E (2009) Sound Localization Cues in the Marmoset Monkey. Hear Res. 260(1-2): 96.

Smith EC, Lewicki MS (2006) Efficient auditory coding. Nature 439:978-982.

Winer JA, Schreiner CE (Eds.) (2005) The Inferior Colliculus. New York, NY: Springer.

Yates GK (1990) Basilar membrane nonlinearity and its influence on auditory nerve rate-intensity functions. Hear Res 50(1-2):145-162.

Young ED (2010) Level and Spectrum. In: The Oxford Handbook of Auditory Science Bethesda, MD: Oxford University Press.

Yu JJ, Young ED (2013) Frequency response areas in the inferior colliculus:nonlinearity and binaural interaction. Frontiers in Neural Circuits 7:90.

Zilany MS, Bruce IC, Nelson PC, Carney LH (2009) A phenomenological model of the synapse between the inner hair cell and auditory nerve: long-term adaptation with power-law dynamics. J Acoust Soc Am 126(5):2390-412.

Biography and Curriculum Vitae

Pre-academic life:

Birth: 11/13/1989, in Martinez, California

High School: Acalanes

Past Academic and Research Experience:

Undergraduate in Biological Systems Engineering at UC Davis, finished in 4 years. Coursework: 3.89 GPA

Took six “advanced” quarters as alternatives to the standard curriculum. These covered more of the fundamental theory behind each topic as well as the standard engineering material.

Internship under Michael J. McCarthy: Analyzed MRI images of tomatoes. Worked with/learned Matlab, realized that it was more than a mathematical package and could be used to automatically crawl through files and folders to extract data for batch analysis.

Internship under R. Paul Singh: Learned/worked with flash, creating “virtual labs” (a teaching tool). Used ImageJ to design simple image processing tools (particle detection). Co-Author of a paper which was awarded “feature article“ for the 11/2013 issue of the Journal of Food Science (DOI:10.1111/1750-3841.12228).

Internship under Shrini Upadhyaya: Constructed DEM simulations (using API of the PFC3D package); main author of a DEM-based paper (DOI:10.13031/2013.39316).

Current Status:

This thesis concludes my biomedical engineering master's (3.93GPA) at Johns Hopkins.