

Spectro-temporal templates unify the pitch percepts of resolved and unresolved harmonics

Shihab Shamma, and Kelsey Dutta

Citation: [The Journal of the Acoustical Society of America](#) **145**, 615 (2019); doi: 10.1121/1.5088504

View online: <https://doi.org/10.1121/1.5088504>

View Table of Contents: <https://asa.scitation.org/toc/jas/145/2>

Published by the [Acoustical Society of America](#)

ARTICLES YOU MAY BE INTERESTED IN

[Formant estimation and tracking: A deep learning approach](#)

[The Journal of the Acoustical Society of America](#) **145**, 642 (2019); <https://doi.org/10.1121/1.5088048>

[Error patterns of native and non-native listeners' perception of speech in noise](#)

[The Journal of the Acoustical Society of America](#) **145**, EL129 (2019); <https://doi.org/10.1121/1.5087271>

[Segregation of voices with single or double fundamental frequencies](#)

[The Journal of the Acoustical Society of America](#) **145**, 847 (2019); <https://doi.org/10.1121/1.5090107>

[Temporal dynamics and uncertainty in binaural hearing revealed by anticipatory eye movements](#)

[The Journal of the Acoustical Society of America](#) **145**, 676 (2019); <https://doi.org/10.1121/1.5088591>

[Deep convolutional network for animal sound classification and source attribution using dual audio recordings](#)

[The Journal of the Acoustical Society of America](#) **145**, 654 (2019); <https://doi.org/10.1121/1.5087827>

[Singing together: Pitch accuracy and interaction in unaccompanied unison and duet singing](#)

[The Journal of the Acoustical Society of America](#) **145**, 663 (2019); <https://doi.org/10.1121/1.5087817>



Spectro-temporal templates unify the pitch percepts of resolved and unresolved harmonics

Shihab Shamma^{a)} and Kelsey Dutta

Department of Electrical and Computer Engineering & Institute for Systems Research, University of Maryland, College Park, Maryland 20742, USA

(Received 9 March 2018; revised 7 December 2018; accepted 9 January 2019; published online 1 February 2019)

Pitch is a fundamental attribute in auditory perception involved in source identification and segregation, music, and speech understanding. Pitch percepts are intimately related to harmonic resolvability of sound. When harmonics are well-resolved, the induced pitch is usually salient and precise, and several models relying on autocorrelations or harmonic spectral templates can account for these percepts. However, when harmonics are not completely resolved, the pitch percept becomes less salient, poorly discriminated, with upper range limited to a few hundred hertz, and spectral templates fail to convey percept since only temporal cues are available. Here, a biologically-motivated model is presented that combines spectral and temporal cues to account for both percepts. The model explains how temporal analysis to estimate the pitch of the unresolved harmonics is performed by bandpass filters implemented by resonances in dendritic trees of neurons in the early auditory pathway. It is demonstrated that organizing and exploiting such dendritic tuning can occur spontaneously in response to white noise. This paper then shows how temporal cues of unresolved harmonics may be integrated with spectrally resolved harmonics, creating spectro-temporal harmonic templates for all pitch percepts. Finally, the model extends its account of monaural pitch percepts to pitches evoked by dichotic binaural stimuli.

© 2019 Acoustical Society of America. <https://doi.org/10.1121/1.5088504>

[JB]

Pages: 615–629

I. INTRODUCTION

Pitch is the basic attribute of sound that conveys musical melody and contributes to speaker identity and speech prosody. It also plays a critical role in enabling listeners to organize cluttered soundscapes into their constituent sources.^{1–5} Yet despite decades of intensive psychoacoustic and physiological experimentation, the underlying neural substrates remain essentially unknown, and the necessary computations are at best ambiguous. Part of the difficulty stems from the ambiguity of the pitch percept itself, which is commonly associated with a variety of stimuli ranging from single and complex tones, to rapidly modulated noise.^{6,7}

We focus here on the pitch evoked by harmonic complexes of different fundamental frequencies. The low order harmonics (<10th) in such a complex are normally aurally fully- or partially-resolved, i.e., their responses are mostly segregated into different frequency channels in the cochlear output as illustrated by outputs up to about 2 kHz in Fig. 1(a). Perceptually, the pitch induced by such harmonics is salient, well-discriminated, with a perceived value corresponding to the fundamental of the complex regardless of whether the fundamental and other nearby harmonics are present or absent. When the harmonics are aurally unresolved (>12th), they co-occur within the cochlear bandpass filters and interact producing a “beating” or amplitude modulations that evoke a pitch sensation corresponding to the

modulation rate [channels beyond about 2.4 kHz in Fig. 1(a)]. However, this “unresolved” pitch is less salient, and is poorly discriminated.^{8–12} Many other contrasts exist between these two pitches that reflect the underlying harmonic separation. For example, randomizing the relative phases among the harmonics leaves the resolved-pitch unaffected, whereas it reduces the saliency of the unresolved-pitch as it distorts the envelope of the beating waveforms. Another key observation concerns the frequency range of these pitch percepts. Unresolved pitches are typically limited only to a few hundred hertz,¹³ a limit whose origin is uncertain,^{14–16} whereas the resolved-pitch is perceived up to an order of magnitude higher frequencies (a few kHz). This latter limit coincides well with the assumed limits of phase-locking on the auditory-nerve^{17,18} (exceptions are discussed later^{19,20}).

The differences between the resolved and unresolved pitch percepts have often led to debates and speculations that they may arise from distinct mechanisms that exploit either temporal or spectral cues. Favoring a unified approach, “temporal models” [Fig. 1(b)] focus on the periodic structure of the phase-locked responses to all harmonics, and propose computations that directly estimate them, thus unifying the extraction of the resolved and unresolved pitch percepts. These models (e.g., the Auto-Correlogram^{16,21,22} and others^{14,15,23}) can explain most of the perceptual properties of resolved and unresolved pitch, including the weak saliency of unresolved pitch and its poor resolution, and the limited range of pitch up to a few hundred hertz.¹³

Spectral models [Fig. 1(c)] by contrast exploit the spectral pattern of the resolved harmonics by matching it against

^{a)}Also at: Laboratory for Sensory Perception, Ecole Normale Supérieure & Paris Science and Letters, Paris, France.

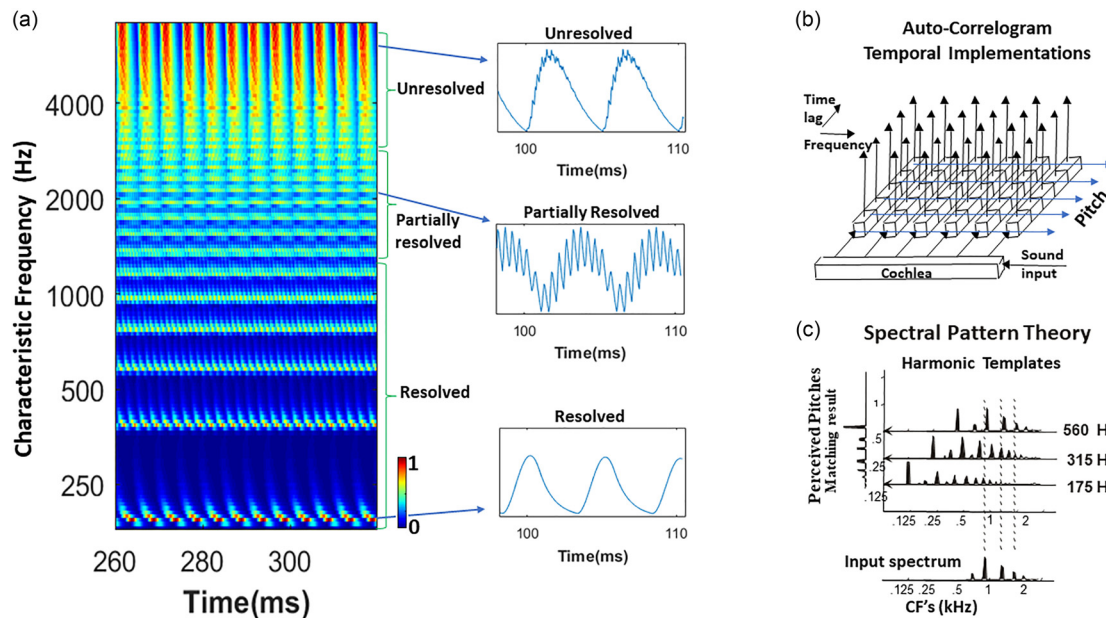


FIG. 1. (Color online) Harmonic resolution and models of pitch perception. (a) Auditory spectrogram of 40 harmonics of 200 Hz. Examples of fully resolved (200 Hz), partially-resolved (2200 Hz), and unresolved (7800 Hz) response waveforms are enlarged and shown to the right. As the harmonics become less resolved, they interact more, causing the response envelopes to “beat” at the difference frequency. (b) The Auto-Correlation model to measure the pitch of both resolved and unresolved harmonics. The autocorrelation function of each cochlear channel response is computed through a chain of temporal delays (lags). The functions are then summed across different channels to produce one combined correlation function whose peaks reflect the common pitch (Ref. 21). (c) Spectral template matching to measure pitch of the resolved harmonics via harmonic templates. The inner-product of the incoming spectrum with each template produces a pattern of matches, whose peaks reflect the perceived pitches (Ref. 25). Color bars next to each spectrogram indicate the scale of the response in arbitrary units.

presumed internal harmonic templates.^{24,25} Since unresolved harmonics lack this spectral signature, these template models do not attempt to account for unresolved pitch, nor for the reasons why the upper limit of resolved-pitch coincides with that of the phase-locking of its harmonics (see Refs. 19 and 20). Consequently, to account for the full range of pitch percepts, spectral models have been supplemented by additional distinct (presumably temporal) mechanisms for extracting unresolved pitch. The two percepts are then unified centrally to deliver consistent percepts.^{26–28}

In a previous study,²⁵ we presented a view that intimately linked temporal and spectral cues in the formation of the harmonic templates. The model demonstrated how temporal phase-locking facilitates the spontaneous formation of the harmonic templates without need for training by explicit harmonic exemplars, and how the templates are subsequently used to estimate the resolved-pitch percept. It, however, did not integrate into these templates the computation of the periodicity from the unresolved harmonics modulations, and therefore left open the fundamental question of how this periodicity can become spontaneously linked to the spectral templates, i.e., how can a 100 Hz modulation rate be heard at the same pitch as that of the harmonic template pattern of a 100 Hz fundamental without explicit training to set up this correspondence?

This report attempts to fill in this gap by proposing a biologically-inspired model that builds upon and augments the harmonic templates model.²⁵ Specifically, our focus is on the unresolved pitch and its unification into the framework of the resolved harmonic templates.²⁵ We shall do this by explaining (1) how unresolved harmonic modulations can

be readily analyzed, and their perceptual values estimated by dendritic processing; (2) how these computations become incorporated into the spectral harmonic templates through the same learning process, which spontaneously gives rise to the templates without need for harmonic exemplars. We shall then (3) discuss why the unresolved pitch range is limited, and (4) implement and simulate the model’s outputs emphasizing stimuli that combine resolved and unresolved harmonics. Finally, we (5) offer a discussion of the physiological evidence of such a model, and ideas for how it might be tested, ending with (6) a brief account of how distinct bin-aural pitches are consistent with this model.

II. METHODS

We summarize here the algorithms used to simulate the function of the proposed extended templates—referred to as the spectro-temporal templates—and how we utilize them to compute the pitch estimates from both resolved and unresolved harmonic inputs.

A. Cochlear frequency analysis and the auditory spectrogram

In many figures in this paper, we display an auditory “spectrogram” representation derived from a computational model of cochlear filtering followed by a lateral inhibitory stage (LIN) that exploits phase-locking in auditory-nerve responses to sharpen the spectral analysis and endow it with robustness to sound-level changes.²⁹ All aspects of this model have already been described in detail in several publications.^{30–33} Briefly, cochlear analysis transforms sound

through a series of stages in the early auditory system from a one-dimensional pressure time waveform to a two-dimensional pattern of neural activity distributed along the tonotopic (logarithmic) frequency axis, referred to as the *auditory spectrogram*. It represents an enhanced and noise-robust estimate of the Fourier-based spectrogram,³² but it differs in many respects due to its biophysical details as previously explained.^{30,31,33}

1. Mathematical formulation

The early auditory stages are illustrated in Fig. 2(a).²⁹ The first operation is an affine wavelet transform of the acoustic signal $s(t)$, which approximates the spectral analysis performed by the cochlear filter bank. This analysis stage is implemented by a bank of 128 overlapping constant- Q ($Q_{10dB} \sim 3$) bandpass filters with center frequencies uniformly distributed along a logarithmic frequency axis (x), over 5.3 oct (24 filters/octave). Each cochlear filter is implemented by a minimum-phase impulse response function $h(t;x)$ with a magnitude frequency response

$$|H(x)| = \begin{cases} (x_h - x)^\alpha e^{-\beta(x_h - x)}, & 0 \leq x \leq x_h, \\ 0, & x > x_h, \end{cases}$$

where x_h is the cutoff frequency, $\alpha=0.3$, and $\beta=8$.³⁴ The cochlear filter outputs $y_{coch}(t,x)$ are transformed into auditory-nerve patterns $y_{AN}(t,x)$ by a hair cell stage consisting of a high-pass filter, a nonlinear compression $g(\cdot)$, and a membrane leakage low-pass filter $w(t)$ accounting for decrease of phase-locking on the auditory nerve beyond about 2 kHz. The final stage simulates the action of a lateral inhibitory network (LIN) postulated to exist in the cochlear

nucleus,³⁵ which effectively enhances the frequency selectivity of the cochlear filter bank.^{31,36} The LIN is approximated by a first-order derivative with respect to the tonotopic axis and followed by a half-wave rectifier to produce $y_{LIN}(t,x)$. The final output of this stage is computed by integrating $y_{LIN}(t,x)$ over a short window, $\mu(t;\tau) = e^{t/\tau}u(t)$, with time constant $\tau = 0.5$ ms mimicking the further loss of phase-locking observed in the midbrain. The mathematical formulation for this model is then summarized as follows:

$$y_{coch}(t,x) = s(t) \otimes_t h(t;x),$$

$$y_{AN} = g(\delta_t y_{coch}(t,x)) \otimes_t w(t),$$

$$y_{LIN} = \max(\delta_x y_{AN}(t,x), 0),$$

$$y_{final}(t,x) = y_{LIN}(t,x) \otimes_t \mu(t;\tau),$$

where $\otimes_t$ denotes convolution operation in the time domain, and δ_t, δ_x are derivatives with respect to time and cochlear axis. The model described above captures several of the important properties of auditory processing that are critical for our objectives, but it is highly simplified and lacks many of the known details of cochlear processing that are likely to be important in certain circumstances and should be added when needed.³⁷ All these issues are discussed in detail in Ref. 29. For MATLAB implementations, see Ref. 38.

B. Models of synaptic plasticity at the spectro-temporal templates

The spectro-temporal templates embody two sets of computations as depicted in Fig. 2(b): (i) bandpass filtering of unresolved harmonics (top panel), and (ii) harmonic pattern matching for the resolved components (bottom panel).

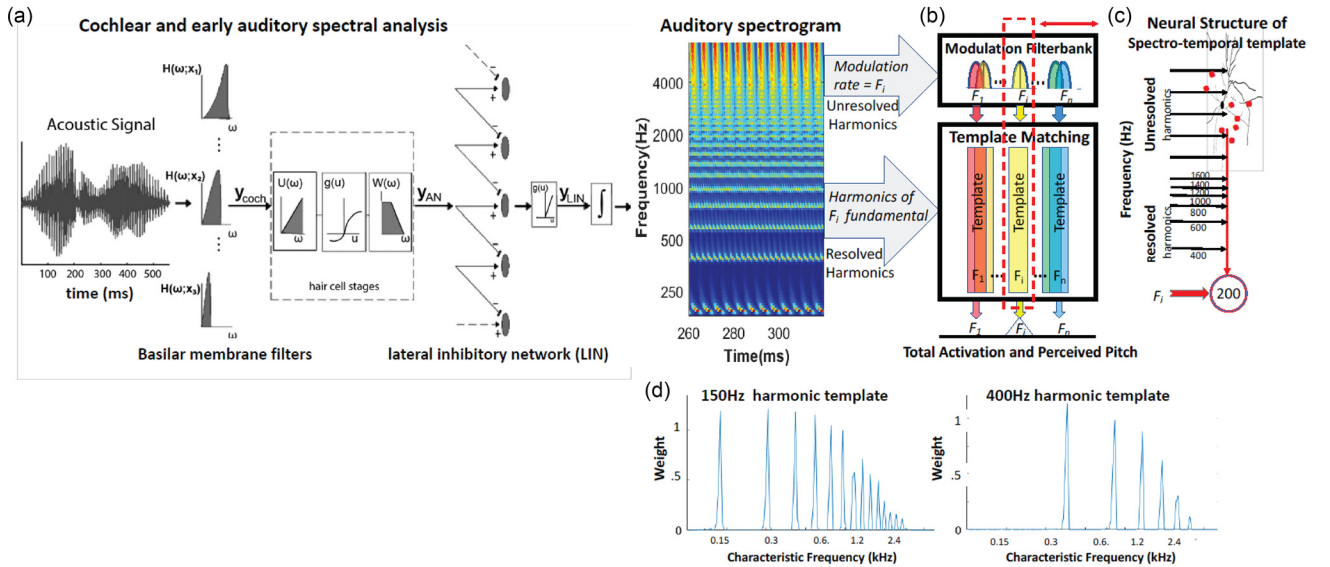


FIG. 2. (Color online) Schematics of the Computational stages in the spectro-temporal harmonics templates model. (a) Schematic of early auditory stages. The acoustic signal is analyzed by a bank of constant- Q cochlear-like filters. The output of each filter (y_{coch}) is processed by a hair cell model (y_{LAN}) followed by a lateral inhibitory network, and is finally rectified (y_{LIN}) and integrated to produce the auditory spectrogram (y_{final}). (b) Schematic of the computations for the unresolved (top panel) and resolved (bottom panel). For the unresolved, periodicity estimation of the amplitude modulated cochlear inputs is done via a bank of bandpass filters. For the resolved components, a harmonic-template matching algorithm is applied (Ref. 25). (c) A proposed biological implementation of the two pitch computations. (d) Two resolved harmonic templates: (left) 150 Hz-fundamental and (right) 400 Hz-fundamental templates. Synaptic weights are applied to each cochlear channel, with peaks at locations harmonic to F_0 .

The full spectro-temporal template is envisaged to be a neuron as depicted in Fig. 2(c); it is referred to later as the “pitch-neuron” for reasons that will be elaborated upon further later in the text. Both resolved and unresolved harmonic inputs from the auditory spectrogram innervate the pitch-neuron and form the appropriate connectivity patterns spontaneously with no supervision from ideal exemplars. The resolved harmonic input synapses are formed as described in detail in a previous publication.²⁵ Thus, we shall focus here on describing the same basic adaptive process that connects the unresolved harmonic inputs into the dendritic region of the pitch-neurons, and results effectively in the formation of a bandpass filter centered at the Characteristic Frequency (CF) of each pitch-neuron.

We begin by assuming that the cochlea is driven by broadband noise. Each pitch-neuron is assumed to be driven by a CF cochlear input that initially induces an intracellular potential that is phase-locked to it. Other contributions to the intracellular potential might be from any other resolved harmonic inputs that have already formed [as schematically depicted in Fig. 2(c)]. These inputs induce phase-locked spiking in the pitch-neuron. The spikes in turn produce intracellular potentials that are phase-locked to them (i.e., at CF), and that backpropagate up the neuron’s dendritic tree. These potentials become attenuated on their way up the dendrites by pathways that are tuned (or resonant) to various frequencies that do not match those of the phase-locked backpropagating signals (at the CF of the neuron). This array of dendritic resonances arises from the anatomical and electrical structure of the dendrites as discussed in the text. Here, it is computationally simulated by a bank of bandpass filters tuned at all frequencies up to about 800 Hz, a range that exceeds the resonance frequencies that are typically found along dendritic locations.³⁹

To simulate synapse formation, we compute the coincidence between the incoming noisy auditory spectrogram and the “backpropagating” dendritic intracellular potentials of the pitch-neurons. For example, Fig. 4(c) depicts for each pitch-neuron (whose CFs are arrayed along the x axis) the strength of these coincidences, which in turn result in formation of synapses (indexed by their dendritic resonances along the y -axis). Thus, the pitch-neuron tuned at CF = 200 Hz [dashed black line in Fig. 4(c)] will form strong dendritic synapses only at locations tuned to 200 Hz, and a few harmonic multiples of that. The same occurs at all other pitch-neurons, a learning process that is elaborated upon further in the text.

C. Monaural pitch estimation

Monaural pitch estimates are made as depicted in Fig. 2(b) from both the resolved harmonics (using harmonic templates) and the unresolved harmonics (via the array of dendritic bandpass filters). For the resolved portion, two examples of resolved harmonic templates [$W^k(f)$] are shown in Fig. 2(d): $W^{150} = 150$ Hz and $W^{400} = 400$ Hz. The templates act a sieve to weight the array of cochlear spectral channels (S_f) at the corresponding harmonic locations. The net output from each template is the waveform

$$O^k(t) = \sum_f S_f(t) W^k(f).$$

Note that this output will be phase-locked to the composite waveform from all inputs. To estimate the overall power in the response, the waveform is half-wave rectified ($g(\cdot)$) and its root-mean-square (rms) power estimated as

$$P_{\text{resolved}}^k = \sqrt{\sum_t g(O^k(t))^2 / T}.$$

Added to the output above is the waveform contribution of the unresolved components, which are computed by filtering each cochlear channel in the unresolved region through bandpass filters centered at pitch k kHz (and a few harmonic multiples $2k, 3k$) representing the CF of each pitch-neuron as depicted in Figs. 2(b) and 4(c). The sum of these inputs ($O_{\text{dend}}^k(t)$) are then added to the resolved-harmonic inputs $O^k(t)$ to form the final pitch-neuron responses, which again are all phase-locked to the sum of the input waveforms. To compute the power in the unresolved pitch responses alone, its contribution is half-wave rectified and then the average power is computed as the

$$P_{\text{unresolved}}^k = \sqrt{\sum_t g(O_{\text{dend}}^k(t))^2 / T}.$$

The total pattern of response power across all pitch-neurons is given by the rms of the sum of the two waveforms

$$P^k = \sqrt{\sum_t (g(O^k(t))^2 / T + g(O_{\text{dend}}^k(t))^2 / T)}.$$

D. Binaural pitch estimation

We propose several possible schemes for binaural interaction of monaural pitch estimates. All approximately produce the results depicted in Figs. 7(a) and 7(c).

Simple summation

$$P_{\text{unresolved}}^k = \sqrt{\sum_t (P_{\text{left}}^k)^2 / T + \sum_t (P_{\text{right}}^k)^2 / T},$$

or simple lateral superior olive (LSO) subtraction and MSO summation [Fig. 7(b), top]

$$P_{\text{LSO}}^k = \sqrt{\sum_t (P_{\text{left}}^k - P_{\text{right}}^k)^2 / T},$$

$$P_{\text{MSO}}^k = \sqrt{\sum_t (P_{\text{left}}^k + P_{\text{right}}^k)^2 / T}.$$

Complex cross-inhibition [Fig. 7(b), bottom]

$$P_1 = P_{\text{est, left}} - P_{\text{est, right}},$$

$$P_2 = P_{\text{est, right}} - P_{\text{est, left}},$$

$$P_{\text{xx}} = \sqrt{P_1^2} + \sqrt{P_2^2}.$$

III. MODULATION RATES DUE TO UNRESOLVED HARMONICS

Speech, music, animal vocalizations, and many environmental and percussive sounds have a broad range of harmonics that extend, in human cochlear analysis, over both the resolved and unresolved ranges. As illustrated in Fig. 1(a), lower resolved harmonics (<10 th) induce localized responses that are phase-locked to their frequencies if less than 3–4 KHz. However, as the resolution deteriorates for the higher harmonics (>10 th), they mutually interact within a cochlear filter, producing envelope modulations at the difference (fundamental) frequency of the complex [200 Hz in Fig. 1(a)]. The depth of this modulation increases as the harmonics become more densely packed. Furthermore, the exact shape of the envelope depends on the relative phases of the interacting harmonics, being strongly peaked if the harmonics are in-phase, and more variable if the phases are relatively random. The carrier of this amplitude-modulated waveform remains phase-locked near the center frequency of the cochlear filter if it is <3 –4 kHz. Beyond this, the carrier is smeared out leaving only the envelope modulations [upper unresolved region in Fig. 1(a)].

As is well-known, subjects listening to the full harmonic complex report a salient pitch that is finely discriminable, and equal to the fundamental of the harmonic series, regardless of its presence or absence.⁶ When listening *only* to the upper unresolved harmonics, subjects still report the same pitch value, but the percept becomes less salient and poorly discriminable. The pitch is weaker still if the relative harmonic-phases are randomized. It is important to appreciate that these envelope modulations can occur at any location along the tonotopic axis where the unresolved harmonics are found. For example, for the 200 Hz harmonic

complex in Fig. 1(a), the 200 Hz envelope modulations become substantial beyond about the 10th harmonic (2 kHz). Consequently, unlike the well-resolved lower harmonics which are found at specific frequency channels (200, 400,... Hz) and hence can be associated with specific fixed harmonic templates, the unresolved pitch is perceived from the envelope modulations *wherever* they occur. This also implies that the same frequency channels (e.g., near 2 kHz) may exhibit envelope modulations at arbitrary rates depending on the unresolved harmonics within the cochlear filter. For example, if the fundamental frequency is <200 Hz, then the channels near 2 kHz would exhibit strong envelope modulations at whatever the fundamental frequency is.

From a computational point of view, a simple scheme to measure these modulation rates, and hence account for the perceived unresolved pitch, is to use a bank of bandpass filters for each cochlear channel [Figs. 2(b) and 3(a)], whose outputs are then summed to generate the overall unresolved pitch value. This scheme is essentially equivalent to the Auto-correlogram model^{21,40} alluded to earlier [Fig. 1(b)], where the delay-lines of the autocorrelation effectively implement the bandpass filters. The same idea is also effectively implemented by the delays proposed in Ref. 25 [Fig. 1(b)]. Physiological experiments have demonstrated rate-selective modulation responses analogous to bandpass filtering,^{41–45} but it remains unclear how these filters exhibit or account for the limited range or poor resolution of the unresolved pitch percepts, and how they can be unified with the resolved pitch percept

A. Dendritic resonances as bandpass filters

Neurons propagate electrical currents from their dendritic trees to the soma, where they induce axonal spiking that

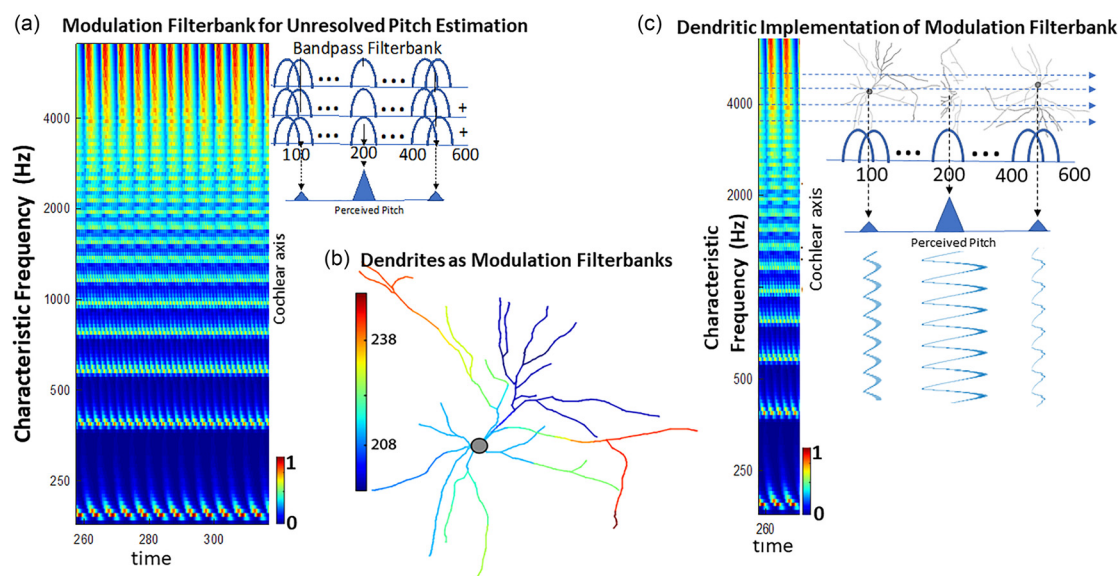


FIG. 3. (Color online) Analysis of temporal modulations on the auditory-nerve. (a) Each auditory-nerve fiber is schematically filtered by a bandpass filter bank tuned to a range of frequencies. The output from all filters is then summed (downward arrows) to estimate the final perceived unresolved-pitch. (b) Schematic of a dendritic tree as a bandpass filter bank (Fig. 4 in Ref. 39). The colors represent the tuning of the path between the synapses on these branches and the soma of the cell (Ref. 54). (c) Implementing the bandpass filter banks with dendritic trees. Each neuron receives inputs from all auditory-nerve fibers whose synapses at a given dendritic tree are at locations that have paths to the soma tuned only near *one* modulation-rate. Thus, there is a neuron whose auditory-nerve inputs are filtered near 100 Hz; others are tuned at 200, 400,... In the schematic, the most responsive dendritic tree is the one tuned to 200 Hz because the auditory-nerve modulations are at 200 Hz. Color bars next to each spectrogram indicate the scale of the response in arbitrary units.

represents the integrated dendritic inputs. Neuronal topology, structure, and active membrane channels have long been known to result in transfer functions between any point on the dendrites and the soma that resemble a bandpass filter, in that it is tuned to a specific frequency that reflects many biophysical parameters. A detailed study of such “bandpass resonances” in a typical dendritic tree [depicted in Fig. 3(b)]³⁹ revealed several important findings that have direct implications for the bandpass filter bank idea needed for unresolved pitch computations. First, in any given dendritic tree [Fig. 3(b)], synaptic inputs at different locations are filtered differently depending on the unique dendritic topology and the details of the path from each synapse to the soma. This means that there exists a range of bandpass filters tuned to different frequencies. Based on a detailed study of such trees,³⁹ it has been found that their basic geometry and structural constraints limit the range of resulting bandpass filter tuning frequencies to a maximum of a few hundred hertz, and their bandwidths to be moderate ($Q = 1$). Therefore, if these dendritic bandpass filters are involved in the measurement of unresolved pitch, their limited range and poor resolution can readily explain the properties of this percept.

Figure 3(c) illustrates schematically how dendritic bandpass filters can serve to estimate the modulation rates induced by unresolved harmonics on the auditory nerve. We first assume that the dendritic trees of a certain type of neurons (we shall refer to as “pitch-neurons”) are initially innervated by auditory inputs from a wide range of cochlear tonotopic locations. To make each pitch-neuron selectively responsive to a single modulation rate only, we shall describe a simple learning (or adaptive) process through which all inputs (synapses) are discarded except those arriving to locations tuned to the same rate. Thus, such a pitch-neuron becomes effectively tuned and responsive only to one modulation-rate. Different pitch-neurons would have dendritic inputs tuned to different modulation rates, effectively creating the bank of bandpass filters needed for unresolved pitch estimation. In Sec. III B, we discuss how this organization naturally emerges without specific training on harmonic exemplars. We shall subsequently address (in Sec. IV) the key question of how these rate-selective dendritic inputs are integrated into the harmonic templates needed to measure the resolved pitch, thus giving rise to the array of pitch-neurons that measure the complete pitch percept.

B. Shaping dendritic rate selectivity

How does the dendritic tree of a pitch-neuron become selective to one input modulation rate? Consider a neuron tuned to a particular CF and possessing an elaborate dendritic tree with auditory inputs from a wide range of cochlear locations [Fig. 4(a)]. The CF is defined by the main input near the soma as indicated in the figure. We assume that the CFs of all such neurons to be $< 3\text{--}4\text{ kHz}$ (upper-limit of phase-locking in most mammals), and hence the CF somatic inputs are phase-locked. Figure 4(a) schematically depicts an array of such neurons with various CFs (200, 350, 700 Hz). We begin with an acoustic noise that excites the cochlea broadly at most of its tonotopic locations. Initially, the somatic CF input at each neuron induces a phase-locked intracellular potential

that excites phase-locked spiking of the cell at the same CF rate (e.g., at 200 Hz), which in turn generates a correspondingly phase-locked signal that backpropagates up to the dendritic tree.⁴⁶ However, this phase-locked CF signal becomes attenuated in the dendritic tree at all branches except those tuned to the same CF. For example, the neuron with $\text{CF} = 200\text{ Hz}$ would phase-lock its spikes to 200 Hz, and the resulting phase-locked intracellular potentials propagate up the dendritic tree, most strongly to dendrites whose pathway to the soma is tuned at 200 Hz.^{47–49} So, if the input to these synapses from the cochlear channels is also phase-locked to 200 Hz, then the synapses may become enhanced because of their potentially correlated pre- or post-synaptic signals. However, if either is weak or unmatched in frequency to the other, then the synapse weakens.^{47–50}

We now consider the spontaneous formation of the rate-selectivity in each recipient neuron when its input is the cochlear responses to *white noise*. At all cochlear CF locations, auditory-nerve responses to the noise are modulated randomly with a bandwidth that reflects the range of frequencies interacting within the cochlear filters [Fig. 4(a)]. We assume initially that the cochlear outputs projecting to the dendritic trees innervate all dendritic locations (indicated by the multicolored synapses in Fig. 4(a)). However, among all these inputs, only those at dendritic locations tuned near the CF of each recipient neuron would have a post-synaptic signal propagating from the soma (e.g., red-synapses for $\text{CF} = 200\text{ Hz}$ neuron, green-synapses for $\text{CF} = 350\text{ Hz}$ neuron, and blue-synapses for $\text{CF} = 700\text{ Hz}$ neuron). Therefore, only these synapses become strengthened by the “Hebbian” correlation between pre- and post-synaptic potentials.⁴⁹ At all other locations, the dendritic post-synaptic signal is too weak and hence the synapses are weakened and eventually pared down. In this manner, we postulate that the array of recipient neurons with different CFs become connected to cochlear inputs that are effectively bandpass filtered with dendritic filters tuned to the same frequency as the CFs. This final arrangement is illustrated for the neuron of $\text{CF} = 200$ in the left panel of Fig. 4(b), where only the red synapses (tuned to 200 Hz) remain, while all others are lost. Such a rate-selective neuron is what we referred to earlier as a “pitch-neuron,” which is tuned only to a specific modulation rate.

We simulated this learning process using a model of cochlear processing that included all major stages such as bandpass filtering, transduction, and phase-locking up to about $3\text{--}4\text{ kHz}$ ^{25,29} (see Sec. II A 1 for more details). Note that this early auditory processing model includes a lateral-inhibitory stage (LIN),^{30,31} which extracts from the phase-locked responses of the auditory nerve a sharp and level-robust spectral representation,^{32,33} equivalent to cochlear tuning of $Q \sim 10$, which has been justified for example in detail in Fig. 2 of Ref. 51. Examples of such auditory spectrograms to noise and harmonic stimuli have also already been illustrated and explained in detail in Refs. 25 and 29. These stages are schematically illustrated in Fig. 2(a).

Similarly, the pitch-neuron and its dendritic tree are modeled by the primary functional parts needed here to explain the learning and filtering process. For example, we

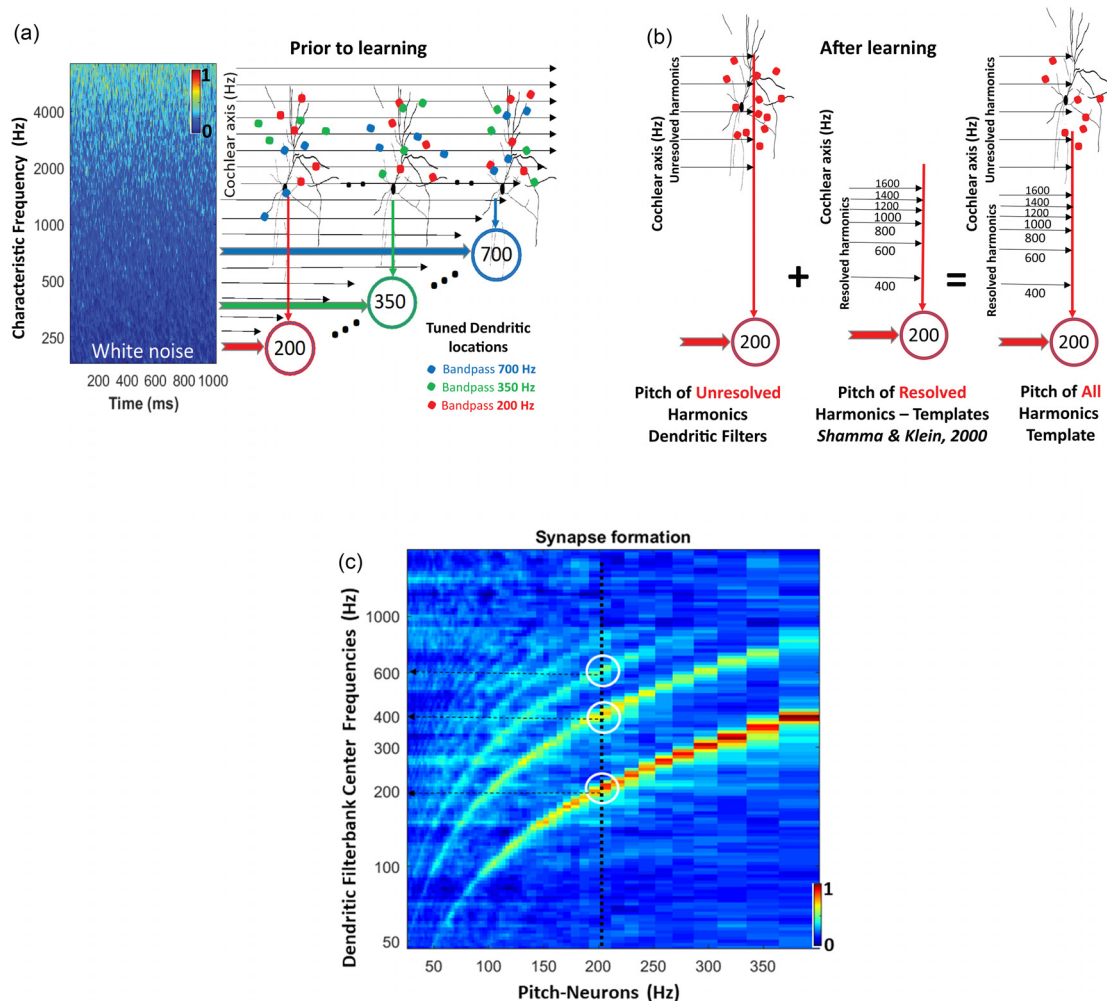


FIG. 4. (Color online) Learning the dendritic bandpass filters. (a) Prior to learning: Schematic of an array of neurons tuned at various CFs. The CF input to the neurons is indicated by the thick colored arrows (shown are CFs = 200, 350, 700 Hz, as red, green, and blue, respectively). Initially each neuron has a dendritic tree innervated by a broad array of cochlear inputs. The synapse locations are symbolized by colored dots that reflect their tuning. Also shown are the cochlear responses evoked by a white noise input. (b) Spectro-temporal templates: (Left panel) Temporal modulations due to unresolved harmonics induce the formation of synapses at dendritic sites that are tuned to the CF of the cell. For example, the pitch-neuron here is tuned at CF = 200 Hz. Only synapses at dendritic locations tuned to the same frequency (=200 Hz) persist, while others are eliminated (compare this to synapses prior to learning). In the end, the sole cochlear amplitude-modulated inputs that excite this neuron are those that are modulated at near its CF rates. (Middle panel) *Resolved harmonic* inputs spontaneously form inputs during the same learning process from inputs at harmonic multiples of the CF, forming the (classic) harmonic template of 200 Hz as described in Ref. 25. (Right panel) Combining all resolved and unresolved harmonic inputs yields the complete pitch-neuron, referred to as the spectro-temporal harmonic template. (c) Simulating the learning: the average coincidence between the intracellular potentials and the noise-generated cochlear responses. It is selectively enhanced only at the dendritic inputs that are tuned at CF (x axis). They connect to auditory inputs from all cochlear locations with filter bandwidths >CF of the neuron (see text). Color bars next to each spectrogram indicate the scale of the response in arbitrary units.

represent the dendritic tree resonances by a bank of bandpass filters with moderate resolution ($Q \sim 1$) as explained earlier based on the dendritic modeling.³⁹ Initially prior to the learning, the somatic signal is the cochlear noisy response at the neuron's CF. This signal causes spiking in the neuron, which in turn induces intracellular potentials that are backpropagated through the dendritic bandpass filters whose outputs represent the post-synaptic potentials available in this neuron.³⁹ Of course, the strongest surviving post-synaptic signal is that at the dendritic locations tuned near the CF [e.g., 200 Hz in Fig. 4(b)]. To simulate the spontaneous formation and strengthening of the synapses, we computed for each pitch-neuron the integrated (over 2 s window) power in the coincidence between all dendritic post-synaptic potentials in this pitch-neuron and the pre-synaptic auditory spectrogram

responses into the dendritic tree. When both the pre- and post-synaptic potentials contain equal frequencies (e.g., 200 Hz), then the average coincidence increases despite their random relative phases (exactly the same as the coincidences previously described in Ref. 25, which resulted in the selective inputs of the resolved harmonics). Figure 4(c) illustrates for each pitch-neuron (arrayed along the x axis) the amount of coincidence (or synaptic strengthening) at all of its dendritic locations, arrayed along the y-axis and indexed by the resonance frequency of each location. For example, in the neuron with CF = 200 Hz (marked with the dashed black line), the strongest formed synapses are those at dendritic locations tuned to 200 Hz, and weaker ones at the multiples of 200 Hz, all of which are highlighted by the white circles. The strong red trajectory highlights the strongest synapses

formed, namely at the dendritic locations (y -axis) that happen to be resonant at the CF of the corresponding pitch-neuron (x axis).

To summarize, the dendritic resonances [Fig. 3(b)] perform the direct measurement of the unresolved-pitch modulations, much like the bandpass filters hypothesized in the schematics of Figs. 3(a) and 3(c). Thus, after learning and synapse strengthening, *unresolved* components of a harmonic complex at high frequencies generate modulated waveforms at the fundamental (e.g., 200 Hz), which best excite the pitch-neuron tuned at the same frequency (CF = 200 Hz), i.e., as in Fig. 4(b) (left panel). However, since dendritic resonances like those described above (and in detail in Ref. 39) are largely limited to a few hundred Hz,³⁹ then only pitch-neurons tuned to such low CFs can utilize the backpropagated somatic potentials in the dendritic trees to form the appropriately tuned synapses. Thus, although temporal modulations created by unresolved harmonics can exceed a few hundred hertz, these high modulation rates remain unperceived because of the lack of corresponding pitch-neurons that can measure them.

Next, we consider the frequency range in the transition between resolved and unresolved harmonics for each pitch-neuron. Pitch-neurons form resonant synapses with cochlear inputs from a wide range of CFs [Fig. 4(b) left-panel] as long as these input channels can provide the temporally modulated inputs (due to the unresolved harmonics). However, the lower limit of the CFs of these cochlear inputs is dictated by their bandwidths, which become progressively narrower at lower tonotopic locations, effectively providing responses to the resolved harmonics. For example, in Fig. 1(a), assuming roughly an effective cochlear tuning of 10% (i.e., $Q = 10$), cochlear outputs below about CF = 2 kHz (10th harmonic of 200 Hz) produce only weak temporal modulations at 200 Hz since they are more resolved. Thus, cochlear channels below this CF will not form the resonant synapses on the dendritic tree of the 200 Hz pitch-neuron. Instead, this pitch-neuron would form the harmonically resolved synapses (the classic harmonic template) as described in detail in Ref. 25. Therefore, as we discuss in more detail next, we recognize two regions of connectivity from the cochlea to each pitch-neuron as illustrated in Fig. 4(b): the resolved harmonics region²⁵ [Fig. 4(b), middle-panel] and the unresolved harmonic inputs (left-panel). The total combination of these inputs is what we refer to as the spectro-temporal harmonic template [Fig. 4(b), right-panel].

IV. SPECTRO-TEMPORAL TEMPLATES: PHYSIOLOGICAL AND PSYCHOACOUSTIC EVIDENCE

In a previous study, we examined how neurons similar to the pitch-neurons spontaneously form synaptic inputs that are harmonically related to the CF of the neuron.²⁵ To review briefly, the post-synaptic somatic potential due to a half-wave rectified and sharpened CF input can correlate with pre-synaptic inputs arriving at from cochlear channels tuned at multiples of the CF. These template connections arise regardless of the input acoustic stimuli, and even if the cochlea is driven by white noise. This is because the inputs

arriving at CF, and from cochlear filters tuned to its low-order multiples (e.g., CF = 200 Hz, and multiples up to approximately the 10th harmonic) are (at least partially) resolved and phase-locked near their carrier frequencies. Consequently, significant coincidence (integrated over a long-duration window) will occur between the somatic CF potential and each of the signals arriving from cochlear filters at multiples of the CF (e.g., 2 CF, 3 CF). As we discussed above, when this signal propagates further up the dendritic tree, higher frequencies are filtered out and the backpropagated signal will only reach regions that are tuned to the CF or its multiples, if they are within a few hundred hertz. Synapses will then form at these locations if the cochlear inputs contain modulations at the CF frequencies, which they do for the noise stimulus discussed earlier [Fig. 4(c)]. After learning, this pitch-neuron becomes connected and driven by the two kinds of cochlear channels [Fig. 4(b)], forming the “spectro-temporal harmonic templates” as we discussed above. Note that in all subsequent figures, each pitch-neuron is now labeled by a pitch value that is both the CF of the neuron (in Hz), or also the fundamental of the harmonic complex that evokes the best response among all pitch-neurons.

In order to illustrate how these pitch-neurons extract and convey the pitch percept, we illustrate in Fig. 5(a) the average power in the responses of an array of such pitch-neurons tuned to a wide range of CFs, and forming an axis which would display a maximum indicating the pitch value by its location, and pitch saliency by its power. The resolved harmonics template used in this and all similar subsequent computations of the resolved harmonics pitch is shown in Fig. 2(d). The stimulus here is composed of a harmonic complex of 200 Hz. The output at each pitch-neuron is computed as detailed in Methods (Sec. II); it consists of the superposition of the inputs coming into the neuron from the resolved (blue traces) and unresolved (green traces) inputs; their superposition is the total activation of the pitch-neurons (red traces). As expected, these activation patterns typically exhibit multiple other peaks that reflect the partial correspondence between different harmonic series. For example, 100 and 200 Hz harmonic series share many harmonics in common. These “simultaneous” peaks in Fig. 5 can be interpreted as concurrently heard pitches whose percept is largely fused with the main pitch (200 Hz).

Figure 5(b) illustrates the decrease in saliency of the unresolved-pitch when the harmonics change from in-phase [Fig. 5(a)] to random relative phases [Fig. 5(b)] causing the peak of the green trace to decrease in height (decrease in saliency). The resolved pitch percept is robust to phase-changes, and here it remains unaltered between the two conditions (blue trace).

Finally, since pitch-neurons are driven by phase-locked cochlear responses both for the resolved and the unresolved harmonics, consequently, the sum of all these harmonic inputs is also phase-locked reflecting all these contributions. Figure 5(c) illustrates the composite time-waveforms of the activation in each of the pitch-neurons (y -axis), all driven by a 200 Hz harmonic complex stimulus. Note that the dominant phase-locking at each pitch-neuron reflects the stimulus

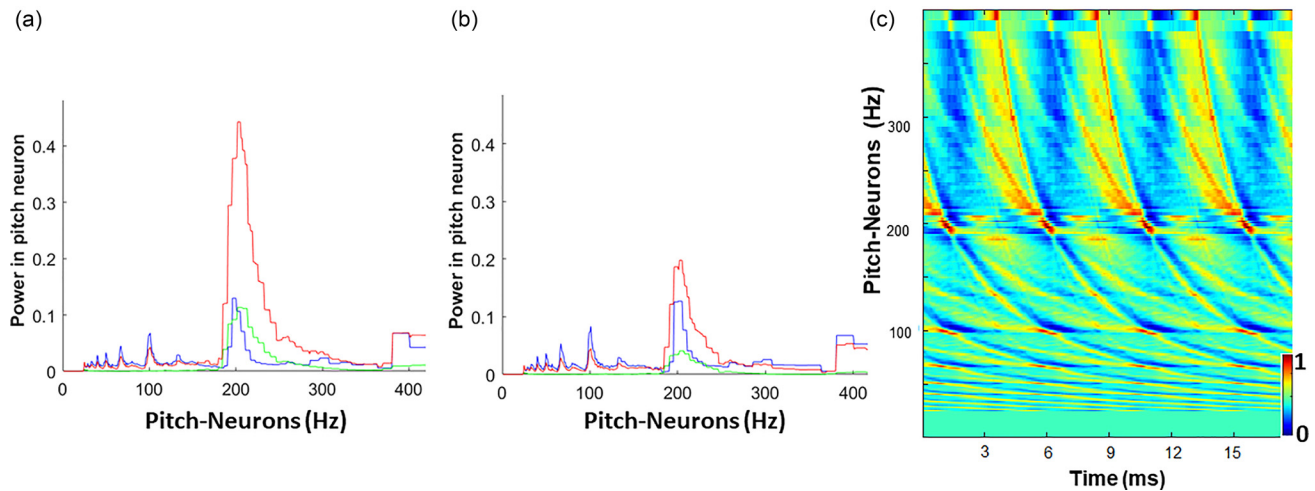


FIG. 5. (Color online) Contributions of the resolved and unresolved harmonics to the outputs of pitch-neurons. (a) Responses of an array of pitch-neurons with CFs (or fundamental pitches) ranging up to 400 Hz. The input stimulus is the 15 component harmonic series of 200 Hz. The resulting activation pattern across the array is computed separately for the resolved (blue) using the spectral templates, and unresolved (green) inputs using the dendritic bandpass filters; the combined activation is also shown (red). The pattern here indicates that the maximally activated pitch-neuron is at 200 Hz. (b) Pitch-neuron array outputs for random phase harmonics. The resolved peak height is unaffected (blue trace), while the peak of the unresolved outputs (green traces) is reduced reflecting its decreased saliency. (c) Output temporal waveforms generated in the same array of pitch-neurons whose average outputs are shown in (a) and (b) above. Color bars next to each spectrogram indicate the scale of the response in arbitrary units.

200 Hz fundamental (5 ms period) and its harmonics that match best each template. For example, the 100 Hz pitch-neuron is phase-locked to 200 Hz, reflecting the strong contributions of the matched harmonics (200, 400,... Hz). Interestingly, scrambling the harmonic phases would change the waveforms even from the resolved inputs, although the location of the peak activations [Figs. 5(a) and 4(b)] remains unchanged.

A. Physiological responses of pitch-neurons

What is the physiological evidence for the existence of such neurons early in the auditory system? As discussed earlier,²⁵ the structure of the postulated pitch-neurons must exhibit dendritic inputs from a wide region across the tonotopic axis, so as to integrate the harmonics of the CF. Such neurons must exhibit broad tuning with multiple harmonically related peaks and are likely also to have high-thresholds for single tones that decrease as more inputs (harmonic tones) are added. The responses of a pitch-neuron are expected to be phase-locked to the composite of its inputs, but most importantly to its CF. Consequently, they are likely to exhibit well-timed onset responses when the inputs are at zero relative-phase. Finally, since phase-locking is essential for the formation of the input synapses, these neurons must have low frequency CFs (<2 kHz) allowing for at least one (phase-locked) harmonic to contribute to the maximum pitch.

In the physiological search and testing of these cells' responses, it is usually very difficult to attain views that match the patterns of Fig. 5(c) because it requires recordings from more than a few pitch-neurons. Instead, experiments typically record from a single neuron, and change the parameters of the stimulus to explore the unit's sensitivity. Thus, if one is to sweep the fundamental frequency of the harmonic complex, the pattern of responses of the hypothetical pitch-

neuron might resemble that in Fig. 6, both for the average activation (left panels) and the phase-locked responses (right panels). Figure 6(a) displays the responses of a pitch-neuron (CF = 400 Hz) to a simple five-harmonic complex as its fundamental frequency (F_0) is swept from 50 to 600 Hz. As expected, it is excited whenever one of the five harmonics falls within its tuning, hence producing peak responses when the $F_0 = 80, 100, 133, 200, 400$ Hz. Note that the response waveforms (right panel) are all phase-locked to 400 Hz because the neuron is always excited by whichever harmonic lies near its CF (400 Hz) and its multiples. This simple response pattern becomes far more complex if the harmonic series contains a substantial number of unresolved components. This is demonstrated in Fig. 6(b) for a 200 Hz pitch-neuron driven by the 4th–15th harmonics of a sweeping complex. The response here reflects both the fine structure of the resolved harmonics, as well as the unresolved modulations. We emphasize, however, that this “physiological view” of one pitch-neuron driven by a sweeping F_0 [Fig. 6(b)] is not equivalent to the view one obtains from the activation pattern across the array of all pitch-neurons as in Fig. 5. Consequently, the peaks in Fig. 6(b) patterns are not necessarily indicative of the perceived pitch since the maximum response may occur at lower F_0 's than the CF of the pitch-neuron. For instance, in Fig. 6(b) the pitch-neuron (CF = 200 Hz) is strongly driven by the $F_0 = 50$ Hz complex because its harmonics happen to coincide with the CF. Nevertheless, this F_0 complex would activate the pitch-neuron at 50 Hz far more strongly, and hence if one is to view the activation induced by this complex ($F_0 = 50$ Hz) across all pitch neurons, one would see a clear peak at CF = 50 Hz.

A different set of physiological experiments in search of the elusive pitch-neurons are those that demonstrated tuning to specific amplitude modulation (AM) rates, mimicking the dendritic filter banks described earlier (Fig. 3). These tuned

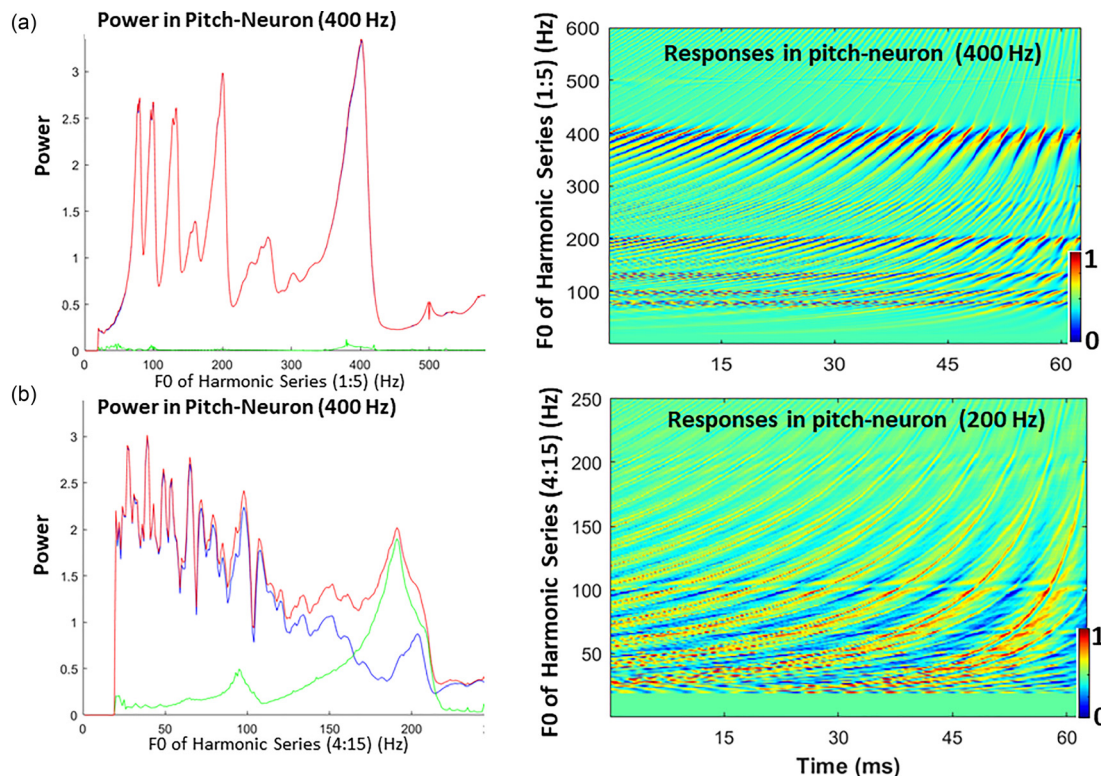


FIG. 6. (Color online) Physiological responses of single pitch-neurons tuned at different CFs. (a) Sweeping a harmonic complex (1:5 harmonics) over a range of fundamental frequencies ($50 < F_0 < 600$ Hz). (Left panel) The pattern of responses in a single pitch-neuron (CF = 400 Hz) is strongly responsive at pitches that are at or multiples of its CF. (Right panel) The waveforms are phase-locked to the CF and occur whenever the harmonics lie within its CF tuning. (b) Phase-locked responses of a pitch-neuron at CF = 200 to a partially resolved complex (4th–15th harmonics). Note the unresolved responses (green trace) are weak when the F_0 is below the CF of the cell (200 Hz). At those low F_0 's, the resolved harmonics contribute but the patterns of response and phase-locking are a complicated mix of the fine structure and envelope modulations. Color bars next to each spectrogram indicate the scale of the response in arbitrary units.

responses, however, have been most commonly described in the inferior colliculus^{42,43} and have not been systematically tested with harmonic tones to demonstrate their relationship to the harmonic templates. Furthermore, the CFs of these neurons are often quite high (>3 – 4 kHz), and their bandpass tuning is broad ($Q = 1$ – 2), suggesting that they may at best serve to estimate the unresolved-pitch. However, the modulation tuning range described (>1 – 2 kHz) is not commensurate with the observed psychoacoustic limits of about 500–700 Hz for the unresolved pitch.¹³ These conflicting data, together with the absence of any information as to how this modulation tuning arises, or its relevance to other pitch properties, cast substantial uncertainty on this physiological evidence.

Finally, it is worth mentioning that maps and responses sensitive to the pitch of the resolved harmonics have been reported in monkey and human primary auditory cortex.^{28,52,53} These findings, however, do not shed light on the mechanisms that give rise to them, because the resolved-pitch percept must arise earlier in the auditory pathway where phase-locking to the harmonic tones is still present, e.g., somewhere before the inferior colliculus.

B. Pitch of binaural stimuli

A harmonic complex is generally heard at the same pitch when presented monaurally or binaurally. However, it has been discovered that the two percepts diverge if the stimuli presented to the two ears are different. Two classic

psychoacoustic studies are discussed here. The first is the Huggins pitch illusion^{54,55} in which two spectrally broad noise stimuli are identical except for interaural π phase-shifts introduced at one or multiple very narrow frequency bands ($1/6$ octaves). The two stimuli are heard as identical noises monaurally; however, the binaural percept contains additional faint tones at the locations of the interaural phase-shifts. Interestingly, when resolved and harmonically spaced, these tones behave just like harmonic complexes, evoking a pitch percept, even if the fundamental of the series is missing. Since such interaural phase-shifts are usually extracted through binaural convergence^{56,57} in the nuclei of the superior olivary complex (SOC), it has then been commonly concluded that the harmonic templates must be located *after* or *more central* to the point of convergence (e.g., at or beyond the SOC complex).

A similar conclusion has been drawn from a second set of experiments in which a harmonic complex is broken up into two sets of components. When presented dichotically, the pitch percept elicited is that of the original complete set. For example, if odd and even harmonics of a 200 Hz fundamental are presented separately to the two ears,¹⁰ the monaural pitches perceived reflect the specific complexes presented to each ear (e.g., 400 Hz in the even-ear, and approximately a 200 Hz pitch in the odd-ear). However, when presented simultaneously, the harmonics are perceived at the pitch of the original complete 200 Hz complex. This reliable finding has also been cited as evidence that the pitch templates used

to estimate the pitch must occur at, or subsequent to the binaural convergence.

However, these conclusions are unwarranted as we demonstrate in Fig. 7. We assume that the spectro-temporal templates are located early along the monaural pathway, prior to the SOC. A key observation to make about the responses of the model pitch-neurons is that they are themselves phase-locked to the combined waveform of all of their (resolved and unresolved) inputs, as was demonstrated in Figs. 5(c) and 6. So, we consider first the Huggins pitch illusion elicited by inter-aural phase-shifts at five harmonic frequencies, e.g., 2nd–6th harmonics of the fundamental 125 Hz (250, 375, 500, 625, 750 Hz). These become evident if we simply subtract the auditory spectrograms of the noise in the two ears [Fig. 7(a)], and obviously applying this pattern to the templates would estimate it correctly as perceived. However, because of approximate linearity and preservation of the phase-locking, this sequence of operations (subtraction followed by the template inner product) can be readily switched while maintaining the same outputs, i.e., $\langle L - R, T \rangle = \langle L, T \rangle - \langle R, T \rangle$. The subtraction then can be done *after* the templates, for example in the SOC, or via the inhibitory cross-projections in the LLN or IC.^{58,59} We emphasize again that all this requires that adequate phase-locking is still preserved at the outputs of the pitch-neurons so as to represent the phase-shifts between the binaural noise stimuli. The schematics in Fig. 7(b) illustrate two possible versions of these computations, and we provide in Sec. II the mathematical formulations implemented to compute the results of the example in Fig. 7(a). The upper panel illustrates the *difference* spectrograms between the two ears, and the pitch-

neurons activation (lower panel) induced by this spectrogram, showing a large peak at the missing fundamental pitch of 125 Hz (red arrow), its octave (250 Hz), and other related pitch frequencies (black arrows).

In Fig. 7(c), we compute and interpret the outputs due to the binaural odd-even harmonic stimuli described earlier. The simplest (if not quite accurate) way to intuit the results is through the linear operations mentioned earlier. Thus, when the monaural patterns are applied separately to the templates, and the outputs added, we get the expected percept described by the experiments and illustrated in Fig. 7(c). Again, this “addition” can occur in the MSO, or even in the identical neural circuits postulated in Fig. 7(b). In fact, the outputs in Fig. 7(c) are all computed using the identical formulations used for the Huggins pitch examples and provided in Sec. II.

V. DISCUSSION

We have described a biologically inspired model for how pitch percepts can be computed via harmonic templates which utilize both temporal and spectral structure of harmonic sound stimuli. The model explains how consistent pitches are estimated from resolved and unresolved harmonics, and how they are subsequently fused to give a unified percept. The postulated pitch-selective template neurons receive synaptic inputs from all cochlear regions. Specifically, they receive cochlear inputs conveying the low-order *resolved* harmonics, forming the classic harmonic templates. They additionally receive the temporal modulations due to the interactions among the higher-order unresolved harmonics, which are subsequently

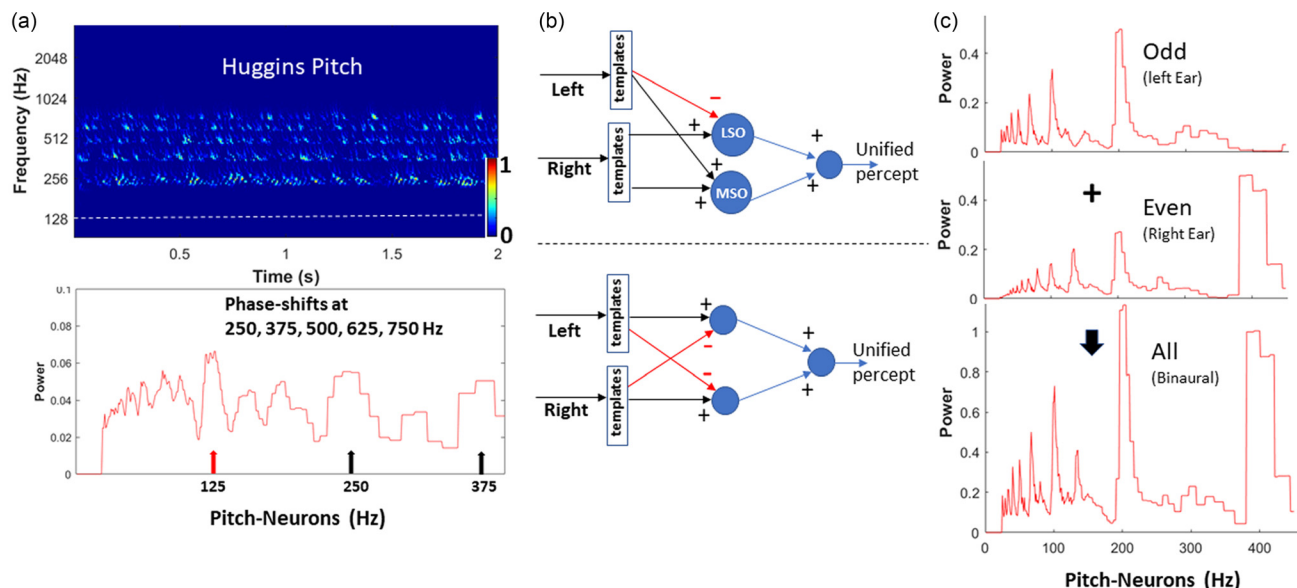


FIG. 7. (Color online) Binaural pitch phenomena do not require *post-binaural* pitch templates. (a) Huggins percept pitch is evoked by highly localized inter-aural phase-shifts, which produce a percept of a tone. These can also be combined with other shifts at harmonic intervals to evoke resolved-pitch percepts and the missing fundamental. (Bottom Panel) Activation in an array of pitch-neurons indicates the percept of the missing fundamental (red arrow) by the Huggins tones (black arrows) at 2nd–6th harmonics of 125 Hz. (b) Schematic of the computations implemented to generate the binaural percepts of the Huggins pitch and the dichotic odd-even harmonics in next panel. (c) Dichotic asymmetric harmonic complexes. Arrays of pitch-neurons are activated by different monaural harmonics (odd and even of 1:10 harmonics of $F_0 = 200$ Hz, presented in separate ears). Each combination evokes the expected percepts (peaks in the pattern for Odd = 200 Hz; and for Even = 400 Hz) when presented separately (top and middle panels). When presented simultaneously (bottom), the percept is that of the full complex (200 Hz). Color bars next to each spectrogram indicate the scale of the response in arbitrary units.

filtered by the dendritic trees of the target neurons, each according to its pitch-selectivity. We have also demonstrated how during development, each pitch-neuron can form all the necessary harmonic-template synapses, and also “tune” its dendritic filter bank through “Hebbian” modification of its synaptic weights. The process can be driven entirely by white noise, with no need for specific harmonic exemplars, and guided solely by the basic tonotopic input into each pitch-neuron.

Pitch estimates through such spectro-temporal templates reproduce the well-known properties of the pitch of resolved harmonics, including the percept of the missing fundamental, and its insensitivity to the harmonic phases. Simulations also show that the harmonic templates need not be situated at or post the binaural structures as has been assumed to explain a variety of dichotic binaural percepts. Instead, we show that monaural templates can readily reproduce the binaural percepts if their phase-locked outputs are combined (summed or subtracted) afterwards.

However, the key new contribution of this spectro-temporal template is its integration of unresolved (temporal) pitch into the classic harmonic (spectral) templates via the tuned dendritic filter banks. The existence of these dendritic resonances is almost unavoidable by the very nature and construction of dendritic trees. The resonances modeled in Laudanski *et al.*³⁹ do not rely on complex precise processes in any way. Just the opposite, they emerge out of basic properties of neuronal membranes and branching patterns as well as basic ionic channel properties. Loosely speaking, the combination of these factors inevitably gives rise to resonances that are fairly broadly tuned and that can account for the decoding of the unresolved pitch as described in the upper branches of the template model. Unfortunately, most research on this topic in dendritic resonances has not looked for or exploited such high frequency resonances. Instead, it has focused on dendritic processing in the classic range of cortical rhythms, which are in the range <10 's Hz. Therefore, the interest in resonances like these is likely to be the exclusive domain of auditory pitch processing! Hence, there is little mention or experimentation to explore this topic beyond what was reported and cited in Laudanski *et al.*³⁹

Finally, the proposed spectro-temporal template elucidates how dendritic filtering imposes a low limit on the resolution and range of the unresolved pitch percept. It also clarifies why temporal modulations at a specific rate are perceived at a pitch equal to that of a tone, or of a harmonic complex with a fundamental, at the same frequency. It should be noted that all these inputs onto the spectro-temporal template or “pitch neuron” are ultimately unified by their harmonic relationship to the CF of the somatic input of the pitch neuron [Fig. 4(b)] because it is this input that determines which resolved harmonics succeed in synapsing onto the cell,²⁵ and which resonant dendritic synapses survive (Fig. 4). However, as we discuss next, our implementation of this unified spectro-temporal template in terms of a “pitch-neuron” model [Fig. 4(b); right panel] is purely speculative. It is indeed entirely possible that other more imaginative and distributed neural circuits could carry out these computations.²⁵

A. Physiological tests of the spectro-temporal templates

The site at which pitch is first computed along the auditory pathway remains one of the fundamental mysteries of auditory neuroscience. It is remarkable that after many decades of neurophysiological experimentation, and with so much detail known about the psychoacoustics of pitch, we still do not know which structures in the early auditory pathway are responsible for the extraction of this crucial percept. Part of the difficulty stems from the myriad variety of temporal and spectral cues available at the cochlear outputs and at subsequent stages which do not decisively exclude numerous contrasting ideas. Another problem is the complex morphology, physiology, and anatomical structure of the cochlear nucleus and following midbrain nuclei, which contain many interconnected subdivisions yielding a bewildering set of pathways up to the inferior colliculus. Finally, the perception of pitch in humans is apparently far more developed than in most mammalian species used in physiological studies of the early auditory system.^{60–62} It is therefore quite possible that the search for the neural substrates of pitch has been frustrated by their absence in these animals!

Nevertheless, assuming that these pitch substrates exist, how can this model guide a new search that benefits from and complements the extensive knowledge that has already been gathered over decades? To start with, there are two clear key requirements that dictate where the templates are likely to form. The first is *temporal*; precise phase-locking is necessary in the formation of the pitch templates. And for many well-understood reasons, phase-locking begins to diminish after the cochlear nucleus (CN), and it is substantially worse by the time we arrive at the inferior colliculus (IC). Consequently, it is likely that the templates form in the CN or just after, but certainly not beyond the IC. The second key requirement is *spectral*; broad convergence of cochlear channels is an inevitable ingredient in the formation of pitch templates. The cells that would respond as pitch-neurons of the model must therefore integrate harmonically related inputs and be phase-locked to the aggregate of these inputs.

There are only a few cell-types in the early auditory pathway, and especially in the CN, that have these properties, and that have been postulated to serve as the harmonic templates. They include the various Onset cells (one of which is also known as the Octopus cell^{58,63–66}) of the Postero-ventral CN (PVCN), which respond robustly to multiple broadly-tuned synchronous inputs and show strong phase-locking to click trains at least up to 2 kHz.^{67,68} The morphology of the octopus cells is also consistent with the requirements of the pitch templates; that it has inputs across the frequency spectrum, and have a cell body located in the lower range of tonotopic organization, and an extensive branching network of dendrites radiating perpendicular to the tonotopic access, thereby accessing the full range of frequency inputs.^{63,64} However Octopus cell projections are excitatory to the ipsilateral ventral nucleus of the lateral lemniscus (VNLL), where their targets provide fast, inhibitory inputs to the contralateral inferior colliculus. These projections do comport with a functional role of onset detection

and reset of neuronal states, but not necessarily for periodicity estimation. Furthermore, it is unclear if the somatic and dendritic inputs to these cells from the auditory nerve are spectrally sharp enough as those illustrated by the auditory spectrograms of Figs. 1–4. In our simulations, the auditory spectrograms are sharpened and rendered level-invariant through the action of a lateral inhibitory network (LIN)^{30,31} (as explained in more detail in Sec. II). The LIN would have to precede the Octopus cells so as to provide the sharpened spectrograms, and no such substrate is known to exist.

Another possible candidate are the neurons of the VNLL, which form extensive and spatially-organized connectivity that has been described as the double-helix of the auditory pathway.^{45,69} The VNLL's primary inputs are the octopus cells in the PVCN and the T-stellate cells of the antero-ventral CN (AVCN). These T-stellate neurons are particularly interesting in that they have been postulated to perform the LIN necessary to extract from cochlear phase-locked responses a robust, sharp spectral ($Q = 10$ – 12), and yet still phase-locked representation regardless of stimulus levels, similar to that generated by the cochlear model in this report and shown in all auditory spectrograms in Figs. 1–4, (see also Sec. II).^{36,70} They have been described as encoding a spectral representation of sound.⁶⁹ VNLL neurons receiving excitatory inputs from T-stellate and octopus cells exhibit multiple morphologies and a large variety of physiological response types whose function remain largely a mystery. Nevertheless, the anatomical organization,^{71,72} the biophysical properties,⁷³ and the few physiological measurements made with harmonically complex stimuli,^{74,75} have inspired proposed schemes of pitch encoding that may be relevant to the model presented.⁴⁵

In summary, it seems that the search for the pitch-neurons in the early auditory pathway may well be productive if it is guided by the findings already available and the general constraints described above. However, for a variety of reasons, there are very few studies of VNLL and CN responses that have directly and simply tested whether the neurons are in fact tuned to harmonic complexes. There are, of course, many difficulties and confounds that can hinder such an investigation and search, including cochlear and neural nonlinearities that render many analysis methods (e.g., Fourier analysis and interval histograms) extremely confusing and distracting. Nevertheless, it is possible to disambiguate many of these measurements and their interpretations with simulations similar to those in Fig. 6.

B. Psychoacoustic challenges of the spectro-temporal templates

As discussed above, much of what we know about the psychoacoustics of pitch perception can be fully accounted for separately by the spectral templates for the resolved pitch, and the temporal filter banks for the unresolved pitch. However, recent investigations have highlighted a challenge to the well-accepted notion that phase-locking is critical for explaining pitch. These concern the finding that pitch perception is possible with resolved harmonics of frequencies well-above the phase-locking range in humans (8–10 kHz). For example, the ability to hear a 2 kHz pitch as the missing

fundamental of a 4th–6th harmonic complex.^{19,20} In our model, the (spontaneous) formation of the spectral templates, and specifically of the synapses of the harmonic inputs, requires that the inputs be phase-locked so as to correlate with the CF of the pitch-neuron. Of course, once the inputs become connected, phase-locking is not essential to convey the harmonic energy.

The simplest way to circumvent the phase-locking limitation above is to postulate that some pitch templates form by repeated exposure to the harmonic complex itself, and not simply by the generalized noise we described before in Ref. 25. Thus, it is commonly found that various patterns of activation (spectral such as vowels or the ultrasonic harmonics of the echolocating Mustache bats⁷⁶) can in principle be learned from examples. In the same way, exposure to the harmonic complex of 2 kHz can link the template (formed already from the low-order harmonics) to the high-order non-phase-locked components when they are presented together. This would explain how these high harmonics of the complex induce the same pitch of the missing-fundamental, and why the resolution of the pitch percept is typical of other resolved pitches.

ACKNOWLEDGMENTS

This work was supported in part by NIH R01DC005779, Advanced ERC grant (Neume), NIDCD training grant No. DC00046, and NSF 1632976.

¹C. Micheyl and A. J. Oxenham, “Pitch, harmonicity and concurrent sound segregation: Psychoacoustical and neurophysiological findings,” *Hear. Res.* **266**(1–2), 36–51 (2010).

²A. D. Patel, *Music, Language, and The Brain* (Oxford University Press, Oxford, UK, 2010).

³D. Schön, C. Magne, and M. Besson, “The music of speech: Music training facilitates pitch processing in both music and language,” *Psychophysiology* **41**(3), 341–349 (2004).

⁴A. S. Bregman, *Auditory Scene Analysis: The Perceptual Organization of Sound* (MIT Press, Cambridge, MA, 1994).

⁵A. J. Oxenham, “Pitch perception and auditory stream segregation: Implications for hearing loss and cochlear implants,” *Trends Amplif.* **12**(4), 316–331 (2008).

⁶B. J. C. Moore, *An Introduction to the Psychology of Hearing* (Brill, Leiden, the Netherlands, 2012).

⁷A. J. Oxenham, “Pitch perception,” *J. Neurosci.* **32**(39), 13335–13338 (2012).

⁸T. M. Shackleton and R. P. Carlyon, “The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination,” *J. Acoust. Soc. Am.* **95**(6), 3529–3540 (1994).

⁹A. J. M. Houtsma and J. Smurzynski, “Pitch identification and discrimination for complex tones with many harmonics,” *J. Acoust. Soc. Am.* **87**(1), 304–310 (1990).

¹⁰J. G. Bernstein and A. J. Oxenham, “Pitch discrimination of diotic and dichotic tone complexes: Harmonic resolvability or harmonic number?,” *J. Acoust. Soc. Am.* **113**(6), 3323–3334 (2003).

¹¹O. Macherey, J. M. Deeks, and R. P. Carlyon, “Extending the limits of place and temporal pitch perception in cochlear implant users,” *J. Assoc. Res. Otolaryngol.* **12**(2), 233–251 (2011).

¹²B. C. J. Moore and H. E. Gockel, “Resolvability of components in complex tones and implications for theories of pitch perception,” *Hear. Res.* **276**(1–2), 88–97 (2011).

¹³O. Macherey and R. P. Carlyon, “Re-examining the upper limit of temporal pitch,” *J. Acoust. Soc. Am.* **136**(6), 3186–3199 (2014).

¹⁴P. Cariani and B. Delgutte, “Neural correlates of the pitch of complex tones. I. Pitch and pitch salience,” *J. Neurophysiol.* **76**(3), 1698–1734 (1996).

- ¹⁵P. Cariani and B. Delgutte, "Neural correlates of the pitch of complex tones. II. Pitch shift, pitch ambiguity, phase-invariance, pitch circularity, and the dominance region for pitch," *J. Neurophysiol.* **76**(3), 1717–1734 (1996).
- ¹⁶R. Meddis and L. P. O'Mard, "Virtual pitch in a computational physiological model," *J. Acoust. Soc. Am.* **120**(6), 3861–3869 (2006).
- ¹⁷J. E. Rose, J. F. Brugge, D. J. Anderson, and J. E. Hind, "Phase-locked response to low-frequency tones in single auditory nerve fibers of the squirrel monkey," *J. Neurophysiol.* **30**, 769–793 (1967).
- ¹⁸D. O. Kim and C. E. Molnar, "A population study of cochlear nerve fibers: Comparison of spatial distributions of average-rate and phase-locking measures of responses to single tones," *J. Neurophysiol.* **42**, 16–30 (1979).
- ¹⁹A. J. Oxenham, C. Micheyl, M. V. Keebler, A. Loper, and S. Santurette, "Pitch perception beyond the traditional existence region of pitch," *Proc. Natl. Acad. Sci. U.S.A.* **108**, 7629–7634 (2011).
- ²⁰B. K. Lau, A. H. Mehta, and A. J. Oxenham, "Superoptimal perceptual integration suggests a place-based representation of pitch at high frequencies," *J. Neurosci.* **37**, 9013–9021 (2017).
- ²¹R. Meddis and L. O'Mard, "A unitary model of pitch perception," *J. Acoust. Soc. Am.* **102**, 1811–1820 (1997).
- ²²M. Slaney and R. F. Lyon, "A perceptual pitch detector," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing ICASSP-90*, Albuquerque, NM (April 3–6, 1990), pp. 357–360.
- ²³J. Laudanski, Y. Zheng, and R. Brette, "A structural theory of pitch: A structural theory of pitch," *eNeuro* **1**(1), ENEURO-0033 (2014).
- ²⁴J. L. Goldstein, "An optimum processor theory for the central formation of the pitch of complex tones," *J. Acoust. Soc. Am.* **54**, 1496–1516 (1973).
- ²⁵S. Shamma and D. Klein, "The case of the missing pitch templates: How harmonic templates emerge in the early auditory system," *J. Acoust. Soc. Am.* **107**, 2631–2644 (2000).
- ²⁶R. P. Carlyon and T. M. Shackleton, "Comparing the fundamental frequencies of resolved and unresolved harmonics: Evidence for two pitch mechanisms?," *J. Acoust. Soc. Am.* **95**, 3541–3554 (1994).
- ²⁷C. J. Plack and A. J. Oxenham, "Overview: The present and future of pitch," in *Pitch* (Springer, New York, 2005), pp. 1–6.
- ²⁸N. Grimault, C. Micheyl, R. P. Carlyon, and L. Collet, "Evidence for two pitch encoding mechanisms using a selective auditory training paradigm," *Percept. Psychophys.* **64**, 189–197 (2002).
- ²⁹T. Chi, P. Ru, and S. A. Shamma, "Multiresolution spectro-temporal analysis of complex sounds," *J. Acoust. Soc. Am.* **118**, 887–906 (2005).
- ³⁰S. A. Shamma, "Speech processing in the auditory system I: The representation of speech sounds in the responses of the auditory nerve," *J. Acoust. Soc. Am.* **78**(5), 1612–1621 (1985).
- ³¹S. A. Shamma, "Speech processing in the auditory system II: Lateral inhibition and the central processing of speech evoked activity in the auditory nerve," *J. Acoust. Soc. Am.* **78**(5), 1622–1632 (1985).
- ³²K. Wang and S. Shamma, "Self-normalization and noise-robustness in early auditory representations," *IEEE Trans. Speech Audio Process.* **2**(3), 421–435 (1994).
- ³³X. Yang, K. Wang, and S. A. Shamma, "Auditory representations of acoustic signals," *IEEE Trans. Inf. Theory* **38**(2), 824–839 (1992).
- ³⁴P. Ru, "Perception-based multi-resolution auditory processing of acoustic signal," Ph.D. thesis, University of Maryland, College Park, MD, 2000.
- ³⁵S. A. Shamma, "Spatial and temporal processing in central auditory networks," in *Methods in Neuronal Modeling*, edited by C. Koch and I. Segev (MIT Press, Cambridge, MA, 1989), pp. 247–289.
- ³⁶R. Lyon and S. Shamma, "Auditory representations of timbre and pitch," in *Auditory Computation* (Springer, New York, 1996), pp. 221–270.
- ³⁷J. R. Cohen, "Application of an auditory model to speech recognition," *J. Acoust. Soc. Am.* **85**, 2623–2633 (1989).
- ³⁸MATLAB implementations of all the above auditory spectrogram computations are available for download at <https://isr.umd.edu/Labs/NSL/Software.htm> (Last viewed 1/22/2019).
- ³⁹J. Laudanski, B. Torben-Nielsen, I. Segev, and S. Shamma, "Spatially distributed dendritic resonance selectively filters synaptic input," *PLoS Comput Biol.* **10**, e1003775 (2014).
- ⁴⁰J. Licklider, "A duplex theory of pitch perception," *J. Acoust. Soc. Am.* **23**(1), 147 (1951).
- ⁴¹G. Langner, "Periodicity coding in the auditory system," *Hear. Res.* **60**, 115–142 (1992).
- ⁴²G. Langner and C. Schreiner, "Periodicity coding in the inferior colliculus of the cat," *J. Neurophysiol.* **60**, 1799–1822 (1988).
- ⁴³A. Rees and A. R. Møller, "Stimulus properties influencing the responses of inferior colliculus neurons to amplitude-modulated sounds," *Hear. Res.* **27**, 129–143 (1987).
- ⁴⁴J. J. Eggermont, "Temporal modulation transfer functions for AM and FM stimuli in cat auditory cortex. Effects of carrier type, modulating waveform and intensity," *Hear. Res.* **74**, 51–66 (1994).
- ⁴⁵G. Langner and C. Benson, *The Neural Code of Pitch and Harmony* (Cambridge University Press, Cambridge, UK, 2015).
- ⁴⁶G. Stuart, N. Spruston, B. Sakmann, and M. Häusser, "Action potential initiation and backpropagation in neurons of the mammalian CNS," *Trends Neurosci.* **20**(3), 125–131 (1997).
- ⁴⁷S. Song, K. D. Miller, and L. F. Abbott, "Competitive Hebbian learning through spike-timing-dependent synaptic plasticity," *Nat. Neurosci.* **3**, 919–926 (2000).
- ⁴⁸R. Kempter, W. Gerstner, and J. L. Van Hemmen, "Hebbian learning and spiking neurons," *Phys. Rev. E* **59**, 4498–4514 (1999).
- ⁴⁹D. O. Hebb, *The Organization of Behavior: A Neuropsychological Theory* (Psychology Press, New York, 2005).
- ⁵⁰H. Markram, J. Lübke, M. Frotscher, and B. Sakmann, "Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs," *Science* **275**, 213–215 (1997).
- ⁵¹R. P. Carlyon and S. Shamma, "An account of monaural phase sensitivity," *J. Acoust. Soc. Am.* **114**(1), 333–348 (2003).
- ⁵²D. Bendor and X. Wang, "Cortical representations of pitch in monkeys and humans," *Curr. Opin. Neurobiol.* **16**, 391–399 (2006).
- ⁵³S. Norman-Haignere, N. Kanwisher, and J. H. McDermott, "Cortical pitch regions in humans respond primarily to resolved harmonics and are located in specific tonotopic regions of anterior auditory cortex," *J. Neurosci.* **33**, 19451–19469 (2013).
- ⁵⁴E. M. Cramer and W. H. Huggins, "Creation of pitch through binaural interaction," *J. Acoust. Soc. Am.* **30**, 413–417 (1958).
- ⁵⁵F. A. Bilsen, "Pronounced binaural pitch phenomenon," *J. Acoust. Soc. Am.* **59**, 467–468 (1976).
- ⁵⁶J. F. Culling, "The existence region of Huggins' pitch," *Hear. Res.* **127**, 143–148 (1999).
- ⁵⁷M. A. Klein and W. M. Hartmann, "Binaural edge pitch," *J. Acoust. Soc. Am.* **70**, 51–61 (1981).
- ⁵⁸J. C. Adams, "Projections from octopus cells of the posteroventral cochlear nucleus to the ventral nucleus of the lateral lemniscus in cat and human," *Audit. Neurosci.* **3**, 335–350 (1997).
- ⁵⁹G. D. Pollak, R. M. Burger, and A. Klug, "Dissecting the circuitry of the auditory system," *Trends Neurosci.* **26**, 33–39 (2003).
- ⁶⁰M. S. Osmanski, X. Song, and X. Wang, "The role of harmonic resolvability in pitch perception in a vocal nonhuman primate, the common marmoset (*Callithrix jacchus*)," *J. Neurosci.* **33**, 9161–9168 (2013).
- ⁶¹K. M. M. Walker, J. W. H. Schnupp, S. M. B. Hart-Schnupp, A. J. King, and J. K. Bizley, "Pitch discrimination by ferrets for simple and complex sounds," *J. Acoust. Soc. Am.* **126**, 1321–1335 (2009).
- ⁶²R. S. Heffner, G. Koay, and H. E. Heffner, "Audiograms of five species of rodents: Implications for the evolution of hearing and the perception of pitch," *Hear. Res.* **157**, 138–152 (2001).
- ⁶³W. S. Rhode, D. Oertel, and P. H. Smith, "Physiological response properties of cells labeled intracellularly with horseradish peroxidase in cat ventral cochlear nucleus," *J. Comput. Neurol.* **213**, 448–473 (1983).
- ⁶⁴P. H. Smith and W. S. Rhode, "Structural and functional properties distinguish two types of multipolar cells in the ventral cochlear nucleus," *J. Comput. Neurol.* **282**, 595–616 (1989).
- ⁶⁵I. M. Winter and A. R. Palmer, "Level dependence of cochlear nucleus onset unit responses and facilitation by second tones or broadband noise," *J. Neurophysiol.* **73**, 141–159 (1995).
- ⁶⁶I. M. Winter, A. R. Palmer, L. Wiegand, and R. D. Patterson, "Temporal coding of the pitch of complex sounds by presumed multipolar cells in the ventral cochlear nucleus," *Speech Commun.* **41**, 135–149 (2003).
- ⁶⁷D. Oertel, R. Bal, S. M. Gardner, P. H. Smith, and P. X. Joris, "Detection of synchrony in the activity of auditory nerve fibers by octopus cells of the mammalian cochlear nucleus," *Proc. Natl. Acad. Sci. U.S.A.* **97**, 11773–11779 (2000).
- ⁶⁸N. L. Golding, D. Robertson, and D. Oertel, "Recordings from slices indicate that octopus cells of the cochlear nucleus detect coincident firing of auditory nerve fibers with temporal precision," *J. Neurosci.* **15**, 3138–3153 (1995).
- ⁶⁹D. Oertel, S. Wright, X. J. Cao, M. Ferragamo, and R. Bal, "The multiple functions of T stellate/multipolar/chopper cells in the ventral cochlear nucleus," *Hear. Res.* **276**(1–2), 61–69 (2011).

- ⁷⁰M. Sachs and C. Blackburn, "Processing of complex sounds in the cochlear nucleus," in *Neurophysiology of Hearing: The Central Auditory System*, edited by R. Altschuler, R. Bobbin, B. Clopton, and D. Hoffman (Raven Press, New York, 1991).
- ⁷¹F. H. Willard and G. F. Martin, "The auditory brainstem nuclei and some of their projections to the inferior colliculus in the North American opossum," *Neuroscience* **10**(4), 1203–1232 (1983).
- ⁷²M. S. Malmierca, T. B. Leergaard, V. M. Bajo, J. G. Bjaalie, and M. A. Merchant, "Anatomic evidence of a three-dimensional mosaic pattern of tonotopic organization in the ventral complex of the lateral lemniscus in cat," *J. Neurosci.* **18**(24), 10603–10618 (1998).
- ⁷³S. H. Wu, "Physiological properties of neurons in the ventral nucleus of the lateral lemniscus of the rat: Intrinsic membrane properties and synaptic responses," *J. Neurophysiol.* **81**(6), 2862–2874 (1999).
- ⁷⁴G. Langner, S. Braun, C. Simonis, C. Benso, and N. Cant, "New evidence for a pitch helix in the ventral nucleus of the lateral lemniscus in the gerbil," in *Proceedings of the Associates for Research in Otolaryngology Abstracts*, Baltimore, MD (February 5–9, 2006), pp. 771.
- ⁷⁵R. Batra, "Responses of neurons in the ventral nucleus of the lateral lemniscus to sinusoidally amplitude modulated tones," *J. Neurophysiol.* **96**(5), 2388–2398 (2006).
- ⁷⁶N. Suga, W. E. O'Neill, and T. Manabe, "Harmonic-sensitive neurons in the auditory cortex of the mustache bat," *Science* **203**, 270–274 (1979).