The case of the missing pitch templates: How harmonic templates emerge in the early auditory system

Shihab Shamma and David Klein
Center for Auditory and Acoustics Research, Institute for Systems Research, Electrical Engineering Department, University of Maryland, College Park, Maryland 20742

(Received 9 April 1999; revised 6 October 1999; accepted 18 January 2000)

Periodicity pitch is the most salient and important of all pitch percepts. Psychoacoustical models of this percept have long postulated the existence of internalized harmonic templates against which incoming resolved spectra can be compared, and pitch determined according to the best matching templates [J. Goldstein, J. Acoust. Soc. Am. **54**, 1496–1516 (1973)]. However, it has been a mystery where and how such harmonic templates can come about. We present here a biologically plausible model for how such templates can form in the early stages of the auditory system. The model demonstrates that *any* broadband stimulus, including noise and random click trains, suffices for generating the templates, and that there is no need for any delay lines, oscillators, or other neural temporal structures. The model consists of two key stages: cochlear filtering followed by coincidence detection. The cochlear stage provides responses analogous to those recorded in the auditory nerve and cochlear nucleus. Specifically, it performs moderately sharp frequency analysis via a filterbank with tonotopically ordered center frequencies (CFs); the rectified and phase-locked filter responses are further enhanced temporally to resemble the synchronized responses of cells in the cochlear nucleus. The second stage is a matrix of coincidence detectors that compute the average pairwise instantaneous correlation (or product) between responses from all CFs across the channels. Model simulations show that for any broadband stimulus, a degree of high coincidence occurs among cochlear channels that are spaced precisely at harmonic intervals. Accumulating coincidences over time results in the formation of harmonic templates for all fundamental frequencies in the phase-locking frequency range. The model accounts for the critical role played by three subtle but important factors in cochlear function: the nonlinear transformations following the filtering stage, the rapid phase shifts of the traveling wave near its resonance, and the spectral resolution of the cochlear filters. Finally, we discuss the physiological correlates and location of such a process and its resulting templates.
[S0001-4966(00)04804-9]

PACS numbers: 43.66.Ba, 43.66.Jh [RVS]

INTRODUCTION AND BACKGROUND

More than any other auditory percept in the last century, pitch has been a potent source of inspiration and controversy in auditory research. Its importance stems from its role in perceiving the prosody of speech, melody of music, and in organizing the acoustic environment into different sources (Summerfield and Assmann, 1990; de Cheveigne *et al.*, 1995). It is generally appreciated that the term ''pitch'' refers to many distinct percepts (de Cheveigne, 1998; Moore, 1989): They include ''spectral pitch'' evoked by sinusoidal signals, ''residue pitch'' (Schouten, 1940; de Boer, 1976) associated with unresolved (high) harmonics, very slow click trains, or the envelope of amplitude modulated noise and sinusoids, and ''periodicity pitch'' (also known as virtual and missing fundamental pitch) evoked by low order, spectrally resolved harmonic tone complexes. The focus of this paper is on ''periodicity pitch,'' the pitch usually associated with musical intervals and melodies, and with speakers voices and speech prosody.

There is general agreement on the perceptual properties and acoustic parameters that give rise to periodicity pitch in humans (and presumably in other mammals and birds) (Langner, 1992; Moore, 1989; Plomp, 1976). For instance, the most salient pitch is evoked by harmonically related tone complexes that are (at least partially) spectrally resolved; the pitch heard is normally that of the fundamental frequency of these harmonics regardless of the energy in that fundamental component; the pitch is roughly in the range 50–2000 Hz. The most effective (or dominant) harmonics are the low order harmonics (the 2nd–5th harmonics). The salience of the pitch increases proportional to the number of resolved harmonics. Multiple pitches are often perceived if there are only a few harmonics in the complex, or if the tones form an inharmonic sequence.

Numerous theories have been proposed to account for periodicity pitch percepts. Most successful among them are the so-called ''spectral pitch theories,'' best exemplified by the ''central pattern recognition'' theories (Goldstein, 1973b; Terhardt, 1974; Bilsen, 1977; Wightman, 1973), and the variations and implementations proposed since then (Duifhuis *et al.*, 1982; Cohen *et al.*, 1995). The two operations common to all are: (1) the pitch value is derived (centrally) from a spectral profile defined along the tonotopic axis of the cochlea (regardless of how this profile is computed); and (2) the input spectrum is compared to internally stored spectral templates, consisting of the harmonic series of all possible

fundamentals.[1] These theories have been enormously successful in explaining and predicting the pitches of complex tones, and consequently have provided the dominant view of pitch perception.

Spectral pitch theories, however, suffer two criticisms. The first is the lack thus far of convincing biological evidence for the existence of these templates or for how they might be generated. ''Learning'' the harmonic templates has usually been assumed to be a straightforward consequence of frequent exposure during early development to speech or natural sounds which tend to be rich in harmonic structure (Terhardt, 1974). However, there are several difficulties with this scenario. Infants are thought to be born with an innate sense of musical pitch (Clarkson and Rogers, 1995; Montgomery and Clarkson, 1997), presumably long before any serious exposure to speech (sounds in the womb are predominantly noiselike due to the heart and other internal organs). Another difficulty is that voiced speech usually has a relatively weaker fundamental component, raising the question of why learned templates consisting of prominent higher harmonics are perceived at (or are linked to) the pitch of the fundamental and not any other arbitrary frequency. A second criticism of spectral pitch theories is their inability to account for other weaker pitch percepts such as ''residue pitch,'' which apparently operate in different parameter ranges, and may require different mechanisms.

To address these criticisms, alternative theories have been proposed to explain how the pitch percept might be computed without the need for stored harmonic templates. These theories can be described as ''temporal'' in that they postulate mechanisms that extract a pitch value from the temporal response in each auditory channel (independent of other channels), and then combine the results from across all channels to get the final estimate. As such, ''temporal'' theories unlike ''spectral'' theories, make no use of an ordered tonotopic axis, i.e., their computations are unaffected by a shuffling of the tonotopic axis (Lyon and Shamma, 1996).[2]

''Temporal'' models vary enormously in the nature of the cues they utilize from each channel, e.g., first or higher order intervals (Evans, 1978; Cariani and Delgutte, 1996a, b; Rhode, 1995; Moore, 1986), autocorrelations of the responses (Slaney and Lyon, 1993; Lickleider, 1951; Meddis and Hewitt, 1991; de Cheveigne, 1998), or synchronization measures and oscillators (Patterson and Holdsworth, 1991; Langner and Schreiner, 1988); they also differ in the mechanisms to measure them, e.g., delay lines and coincidence detectors, or intrinsic oscillators. One often stated advantage of these theories is that most can account for both residue pitch, as well as periodicity pitch, with the same mechanisms.

However, just as with the spectral models, the temporal models suffer from certain shortcomings. For instance, the physiological basis of these models is also uncertain. Thus while many central auditory responses can be interpreted as exhibiting delays or appropriate oscillatory patterns, the anatomical and physiological data do not yet coalesce as a whole into a compelling picture (Langner, 1992). Furthermore, most physiological pitch data tend to be in frequency ranges and from units with best frequencies that are relevant for residue pitch (30–300 Hz) or slow temporal modulations (<30 Hz) (Schreiner and Urbas, 1988; Schreiner and Langner, 1988; Schwartz and Tomlinson, 1990) rather than periodicity pitch. Finally, recent psychoacoustical findings have been interpreted in favor of a dual (rather than a unitary) model of pitch perception (Carlyon, 1998a).

In summary, it is fair to say that spectral pitch theories would be more palatable to many: (1) if there is a biologically compelling mechanism for how harmonic templates might come about, and evidence for their existence; and (2) if the models could be extended to take into account ''residue pitch'' percepts and their properties. This paper addresses primarily the first issue, and provides ideas for what physiological mechanisms and anatomical substrates are potentially involved, and where to search for them. An important goal of this paper is to demonstrate that harmonic templates may emerge as a consequence of basic properties of early auditory processing, and not of exposure to any special sound stimuli such as harmonically rich speech or music. To emphasize this point, we shall use broadband noise and irregular click trains (i.e., sounds that lack any harmonic character) to produce the harmonic templates. We shall also briefly touch upon the problem of residue pitch, and discuss potential candidate mechanisms for unifying the estimation of both periodicity and residue pitch percepts without resort to organized correlation delay lines and other purely temporal structures.

The model we describe here explains how harmonic templates could emerge as a simple consequence of coincidence detection among channels representing the outputs of a cochlearlike filter bank. Once formed, the templates can be used to estimate the pitch as in the many variants of the spectral-matching pitch algorithms. Our focus in this paper is on the template-formation phase. Our goal is to illustrate how biologically plausible processes, response patterns, and connectivity in the early auditory nuclei can give rise to ordered harmonic templates without the need for any specially tailored inputs (such as clean harmonic complex tones), or supervised constraints (such as labeled and ordered inputs and outputs).

In the following, we shall first illustrate the essential mathematical structure of the model (Sec. I) and then discuss why the templates emerge (Sec. II). Next we discuss the potential biological structures and pathways that underlie the model (Sec. III). The implications of this model to the encoding of residue pitch are discussed in Sec. IV. We finally discuss the wider implications of our findings to models of auditory processing and to neural processing strategies in general (Sec. V).

I. A MATHEMATICAL MODEL FOR HARMONIC TEMPLATE GENERATION

The two basic stages of the model are illustrated in Fig. 1: An analysis stage consists of filter bank followed by temporal and spectral sharpening analogous to the processing seen in the cochlea and cochlear nucleus. The second stage is a matrix of coincidence detectors that computes the pairwise instantaneous correlation among all filter outputs.

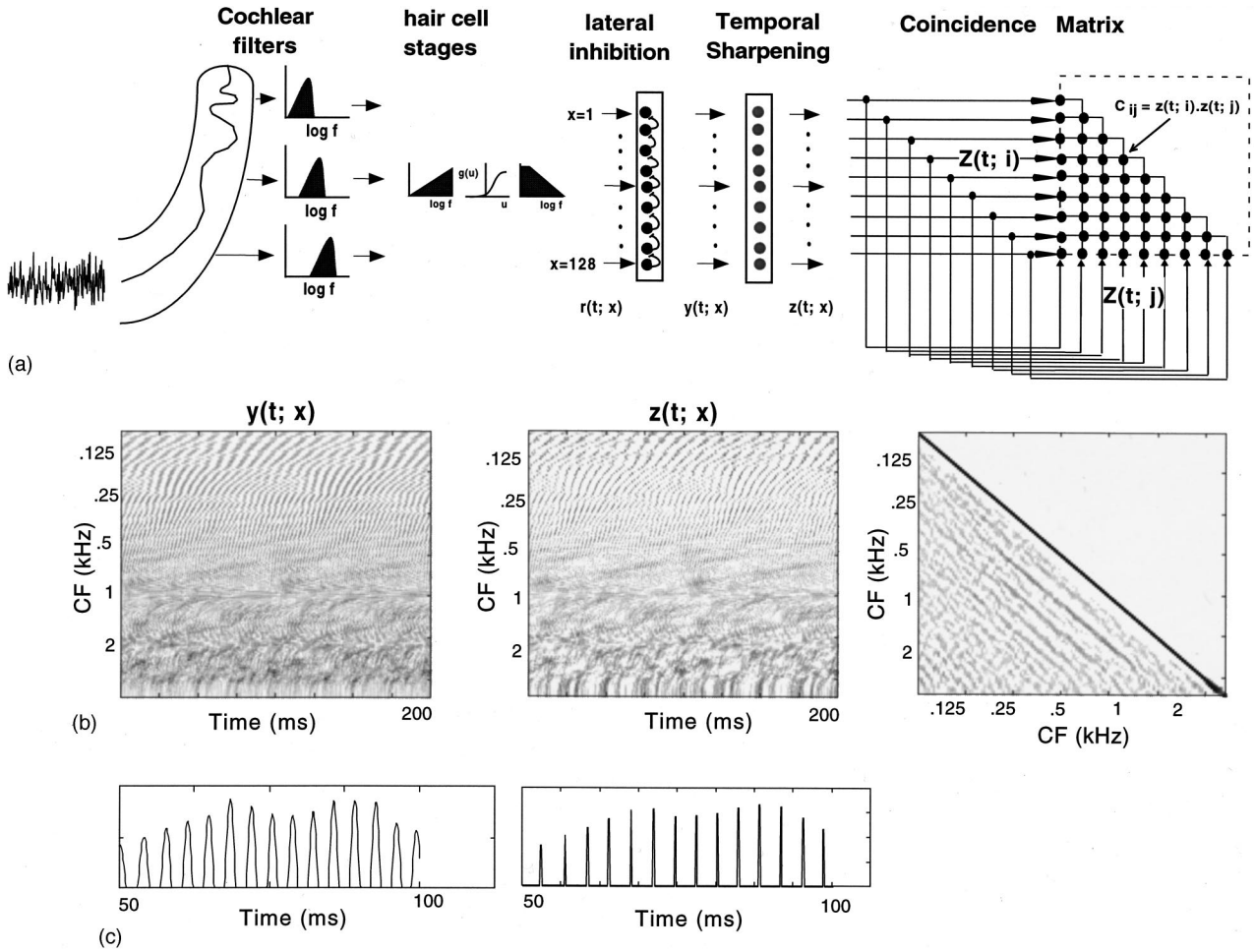

FIG. 1. Schematic model of early auditory stages. (a) Sound is analyzed by a bank of 128 tonotopically ordered cochlear filters spanning CFs between 100 and 4000 Hz. The output waveform from each filter is passed through a hair cell model ($r(t;x)$), followed by a first difference across the channel array simulating the action of a lateral inhibitory network (LIN) ($y(t;x)$). The responses are then temporally sharpened, becoming more synchronized within each channel ($z(t;x)$). The final stage is a matrix of coincidence detectors that compares the responses from all pairs of channels across the array. (b) The spatio-temporal responses of the channel array at different stages of the model: (left-to-right)—The responses at the LIN output ($y(t;x)$); the synchronized responses ($z(t;x)$); the output of the coincidence matrix (C) after one iteration. (c) The waveform transformation at the synchronization stage. (Left)—the waveform at CF$\approx$140 Hz ($y(t;x=12)$). (Right)—the waveform after temporal sharpening.

A. The analysis stage

This stage consists of a simplified minimal model of early auditory processing. It consists of a cochlear filter bank, followed by hair cell rectification and central spectro-temporal sharpening. These operations are depicted in Fig. 1, and described below in detail.

1. Cochlear filter bank

We employ a bank of 128 bandpass filters, equally spaced along a logarithmic frequency axis, x with center frequencies (CF) spanning a range of 5.3 octaves. The filters are moderately tuned and significantly asymmetric, with a steep roll-off on the high-frequency sides, as illustrated in Fig. 2(a) (Wang and Shamma, 1994; Yang *et al.*, 1992). They have constant Q's, and hence their bandwidths gradually broaden (on a linear scale) toward the higher CFs. They are also related to each other by a simple dilation of their impulse responses. Given a discrete-time signal $s(t)$, and cochlear filter impulse responses $h(t;x)$, $x=1,\ldots,128$ and $t=0,1,\ldots,n$, any filter's response is computed as

$$u(t;x)=s(t)*h(t;x), \tag{1}$$

where $*$ denotes convolution with respect to time.

2. Hair cell filtering and rectification

Hair cells convert the filter outputs into electrical activity along the tonotopically ordered auditory-nerve array. This biophysical process is usually modeled by a three-step process (Shamma *et al.*, 1986; Shamma and Morrish, 1986): a high pass filter accounting for the velocity coupling of the hair cell cilia; a sigmoid function that describes nonlinear hair cell transducer channels; and a low pass filter representing the leakage in hair cell currents that gradually attenuates phase-locked responses beyond 800 Hz.

Here we shall simplify the analysis by incorporating the first temporal derivative into the cochlear filters. Next, the hair cell nonlinearity $g(\cdot)$ is modeled as a simple half-wave rectifier:

$$r(t;x)=g(u(t;x))=g(s(t)*h(t;x)), \tag{2}$$

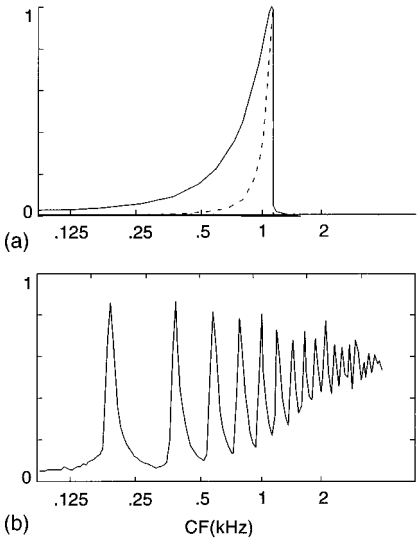

FIG. 2. Details of the model filters. (a) (solid line) Magnitude transfer function of the filter at CF=1 kHz; (dotted line) The effective magnitude transfer function *after* the LIN stage (see text). (b) The integrated output of the LIN ($\Sigma_t|y(t;x)|$) reflecting the spectrum of a harmonic series stimulus consisting of 20 harmonics of a 200-Hz fundamental [see Wang and Shamma (1994) for details].

where $g(u)=0$ for $u<0$, and $g(u)=u$ otherwise. Note that $g(\cdot)$ can be redefined as a sigmodal function to account for more complex nonlinear effects such as saturation or wider dynamic ranges. The effects of these added modifications is small for reasons discussed later. The hair cell low pass filter is bundled into the following stage as we describe next. The model outputs at this stage are depicted in Fig. 1(b)–(c) for a broadband noise stimulus.

3. Spectral and temporal sharpening of the filter outputs

This stage is helpful in enhancing the representation of the harmonics in the templates as we shall discuss later. Spectral sharpening mimics the effect of lateral inhibition (Shamma, 1985a, b), and is modeled by a simple derivative across the channel array (or a first-difference operation between the filter outputs) (Wang and Shamma, 1994):

$$y(t;x)=r(t;x)-r(t;x-1), \tag{3}$$

for $x=2,...,128$, and $y(t;1)=0$. It can be shown that this step effectively sharpens the cochlear filters (Wang and Shamma, 1994; Lyon and Shamma, 1996), and is in principle unnecessary if the cochlear filters used are sharp enough to resolve approximately up to eight harmonics. Figure 2(b) illustrates that our frequency analysis at this stage can partially resolve approximately 8–10 harmonics of a 20-harmonic series stimulus (with a fundamental at 200 Hz).

The next stage performs temporal sharpening which enhances the synchrony of the phase-locked responses. This process mimics transformations such as those seen between the auditory-nerve and the onset units of the cochlear nucleus (Oertel *et al.*, 1990; Palmer *et al.*, 1995; Rhode, 1995). It is approximated by sampling the *positive peaks* of $y(t;x)$:

$$z(t;x)=\sum_{t_p}\delta(t-t_p)\cdot y(t_p;x), \tag{4}$$

where t_p=locations of the positive peaks in time, and $\delta(\cdot)$ is the discrete Dirac delta-function [$\delta(0)=1$ and $\delta(\cdot)=0$ otherwise]. $z(t;x)$ then becomes a spectrally sharpened and highly temporally synchronized version of the filter responses as illustrated in Fig. 1(b)–(c). A simple way to include the effects of the diminishing phase-locked responses with increasing frequency is to replace $\delta(\cdot)$ with a pulse of variable width $\Pi_m(\cdot)$, starting at the zero-crossing point, i.e., $\Pi_m(k)=1$, $0\leqslant k\leqslant m$; the larger m is, the smaller is the frequency range of phase-locking and synchrony [Fig. 1(c)]:

$$z(t;x)=\sum_{t_p}\Pi_m(t-t_p)\cdot y(t_p). \tag{5}$$

B. The coincidence matching stage

This stage performs an instantaneous match between the responses of all pairs of channels in the array, and integrates all results over time to produce its final output. From a mathematical perspective, the network is a matrix of coincidence detectors, each multiplying the responses from a pair of channels as depicted in Fig. 1(a):

$$C_{ij}(t)=z(t;i)\cdot z(t;j), \tag{6}$$

for all $i,j=1,...,128$ such that $j<i$; for $i=j$, $C_{ii}(t)=z(t;i)$. The absolute values of $C_{ij}(t)$ are then accumulated over time until an adequately smoothed output T_{ij} is obtained:

$$T_{ij}=\sum_{N,t}|C_{ij}(t)|, \tag{7}$$

for N realizations of a random stimulus. Note there are no neural delays anywhere in this model. Instead, coincidences are computed from simultaneous outputs of the filter bank, and the results are then integrated over time. Note also that in the equation above, it is the *absolute* value of the coincidence that is integrated, and that the average value of $z(\cdot)$ can be removed because it is about the same for all channels, and hence contributes only a uniform constant at all locations.

C. Model simulations

The coincidence network above is capable of producing the harmonic templates as its final averaged output regardless of the exact nature of its input signal, provided it is broadband conveying energy at all frequencies <3 kHz. We illustrate in Fig. 3 the templates generated with broadband noise and random click train input signals $s(t)$. In Fig. 3(a), the 200-ms stimulus consists of equally spaced, random-phase tones (with 10-Hz separation, in the range between 10 Hz and 4000 Hz) with random phases. Usually many examples of $s(t)$ are generated with different random phases. The final output of the network (T_{ij}) is the average over all these stimulus iterations [$N=300$ in Fig. 3(a)]. Figure 3(b) shows the average output T_{ij} for a random click train stimulus with random widths ($N=300$).

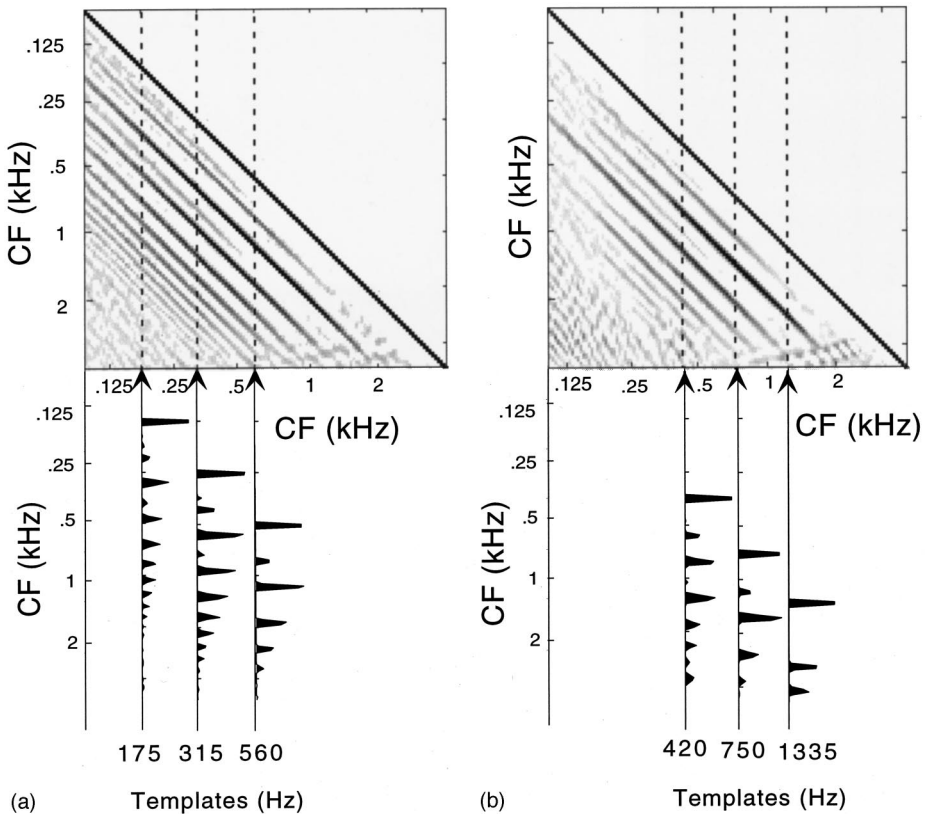

FIG. 3. The harmonic templates in the integrated output of the coincidence matrix. Templates emerge as regions of high coincidence that run parallel to the main diagonal, and are exactly spaced at harmonically related CF distances. (a) The templates generated by a broadband noise stimulus. Three templates are shown individually by the cross sections (fundamentals at 175, 315, 560 Hz). For each, the pattern shows prominent peaks at harmonically related CFs, that gradually decrease in amplitude for higher order harmonics. (b) The templates generated by a random click train with random widths. Cross sections for three templates are shown below the figure (fundamentals at 420, 750, 1335 Hz).

The simulations show strongly correlated outputs from channels that are separated exactly by harmonic distances from each other. These strong coincidences form a pattern of multiple diagonals that are spaced at exactly harmonic intervals apart. For instance, consider the pattern of strong coincidences for the channel at CF=175 Hz displayed below the coincidence matrix outputs in Fig. 3(a). The pattern shows prominent peaks at CFs that are integral multiples of 175 Hz. This pattern is interpreted as the ''harmonic template'' of the 175-Hz series. Similarly, the templates for all other harmonic series can be found across the diagonals of the network output (cf. the harmonic series templates for several other fundamentals in Fig. 3). Note also that the number of harmonics represented in each template decreases with increasing fundamentals as phase-locking diminishes gradually beyond 1 kHz.

D. Final comments

The mathematical structure of the network and simulations described above are but one example of many variants that can be used. The two key operations are a cochlearlike filtering stage followed by coincidence detection. Relaxing the degree of spectral and temporal sharpening in the model only gradually reduces the clarity of the templates by either diminishing the height of the harmonic peaks or reducing their number. Similarly, replacing the ''product'' in the coincidence operation [Eq. (6)] with a squared sum or other ''matching'' operations does not alter the locations of the harmonic peaks. The reasons behind this robustness are discussed in the next section.

II. WHY DO THE HARMONIC TEMPLATES EMERGE?

In this section, we examine the reasons why the coincidences in Fig. 3 occur at harmonic intervals between the cochlear channels despite the lack of any harmonic structure in the input stimuli. We shall specifically discuss the critical role played by three subtle but important factors in the model: the nonlinear transformations following the filtering stage; the rapid phase shifts of the traveling wave near its resonance; and the spectral resolution of the cochlea.

A. Nonlinear transformations of the filter outputs

In the model outputs, the harmonic-template lines emerge as a consequence of the strong coincidences between responses of harmonically related cochlear filters. To understand why this is so, consider a sharply tuned filter bank driven by a broadband noise stimulus. Each filter in this bank produces a phase-locked response waveform that is quasi-periodic and reflects predominantly its CF (Ruggero, 1973). This is exemplified in Fig. 4 by the quasi-sinusoidal responses for a CF$\approx$250 [Fig. 4(a)]. If the filter outputs are not half-wave rectified or otherwise nonlinearly distorted, then the outputs from any such pair of filters will be orthogonal (or entirely uncorrelated) since each would contain Fourier coefficients only near its CF.

However, the situation is drastically different if the filter responses are half-wave rectified, because this creates ''distortion'' components and the waveform can be thought of as composed of a fundamental frequency (the CF of the filter) and its harmonics [Fig. 4(b)]. Consequently, the rectified waveform from any filter can now partially coincide with

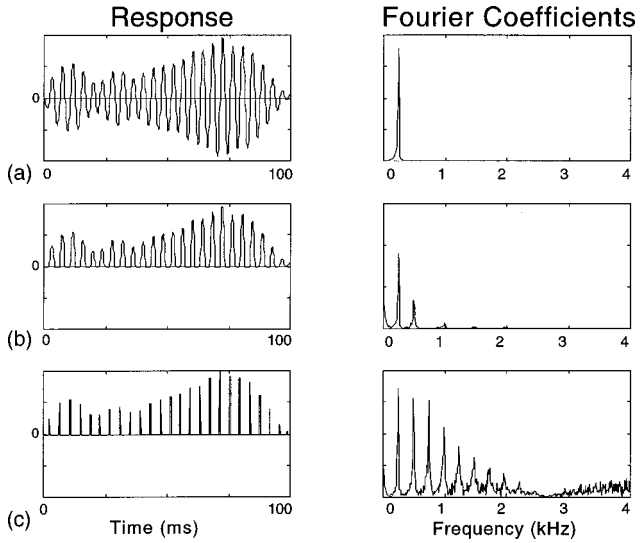

FIG. 4. The effects of nonlinear deformations and temporal sharpening. The response waveforms within each channel implicitly convey harmonic-distortion components with varying strength. The left column illustrates response waveforms with increasing nonlinear distortion and synchrony; the right column illustrates the Fourier coefficients corresponding to each waveform. The number and amplitude of the harmonic distortion components increase with increasing synchrony and nonlinear deformation of the response waveform. (a) Linear filter response at CF $\approx$250 Hz; (b) half-wave rectified response; (c) the synchronized impulse train corresponding to the 250-Hz response.

outputs of other filters that are at harmonically related CFs. For instance, the rectified waveform from the filter at CF $=$250 Hz contains harmonics of 250 Hz with gradually decreasing intensity [Fig. 4(b)], and hence may coincide strongly with filter outputs at CF$=$500,750,... Hz. The important role played by the half-wave rectification is not unique to this operation; rather, it is a common consequence of many instantaneous nonlinear distortions of the filter outputs. For example, similar harmonic coincidence patterns emerge if the filter waveforms are distorted by a saturating nonlinearity, a limited dynamic range, or are converted to a series of synchronized impulses as is done in the model; Eq. (4) [Fig. 4(c)].

It is in this context that one can appreciate the role of enhanced temporal synchrony in the model. The synchronization of the filter response waveforms is a highly nonlinear operation that ensures that the impulse train from each filter contains within it the fundamental frequency (at the CF) and, prominently, many of its harmonics [Fig. 4(c)]. That is why the pulse train from a filter at CF$=$250 Hz will correlate well with pulse trains produced by filters at harmonically related CFs up to a relatively high order.

B. The phase of the cochlear traveling wave

How is it possible that the highly synchronized waveform at a given CF (e.g., 250 Hz) be in just the right phase to coincide with outputs from other CFs (e.g., the response at 500 Hz)? The answer highlights the role of the cochlear traveling wave, specifically its phase delays, in the formation of the templates.

Figure 5 illustrates the typical features of two traveling waves evoked by two tones, say at 250 and 450 Hz. Near the

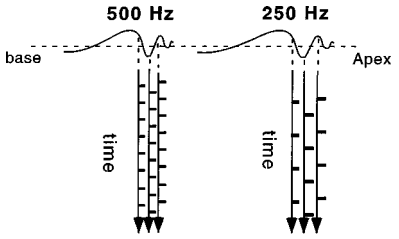

FIG. 5. Traveling wave phase shifts near the resonance of a traveling wave. The schematic illustrates that the response patterns near the resonance of the traveling waves can be significantly phase shifted relative to each other over very short distances.

resonance of each wave (CF$=$250 and 500 Hz), the travel velocity decreases rapidly, and the wave as a result accumulates phase delays at an accelerated pace (Lyon and Shamma, 1996; Shamma, 1985a). Consequently, near the CF, one may find responses of widely different (even opposite) phases in closely spaced locations (or channels). That is, each of the CF regions of 250 and 500 Hz contains synchronized responses to these frequencies at various phases, and hence it is likely that at least a pair of channels will coincide and positively correlate. This argument still applies when the stimulus contains many tones (as with a harmonic complex or broadband noise) because these phase delays are characteristic of the cochlear filters and not of the stimulus. Thus as long as the responses at a given CF are determined by a relatively sharply tuned cochlear filter, they will necessarily exhibit these rapid phase shifts near the CF, as can be seen in Fig. 1(b) where the synchronized responses to the noise stimulus are similar in adjacent channels except for a rapid phase delay toward the lower CFs. Finally, note that the formation of the templates requires only that the local phase shifts be relatively rapid, and is insensitive to changes in absolute phase values or in the detailed shape of the rapid phase functions. Consequently, nonlinear phase changes such as those induced by increasing sound levels and other manipulations are unimportant as long as they leave the travelling wave phase functions relatively rapid near the resonance.

C. The sharpness of frequency analysis

Cochlear frequency analysis and subsequent spectral sharpening of the filter outputs [by lateral inhibition [Shamma, 1985b; Eq. (3)] enhance the features of the harmonic templates. This is because sharp filters (by definition) respond only to frequencies near their center frequencies, and hence usually produce more regular (periodic) synchronized responses regardless of the nature of the input stimulus. This point is illustrated in Fig. 6 where we examine the effect of broadening the cochlear filters on the synchronized responses to a broadband noise stimulus. Figure 6(a) shows the synchronized responses (left plot) and their corresponding Fourier series coefficients (right plot) using our regular filters. Here, the response at each CF contains well-defined components at CF and its harmonic distortions as is evident by the well separated Fourier peaks. If the filters are made significantly broader (for instance, by removing the lateral inhibition stage), the synchronized responses from each filter

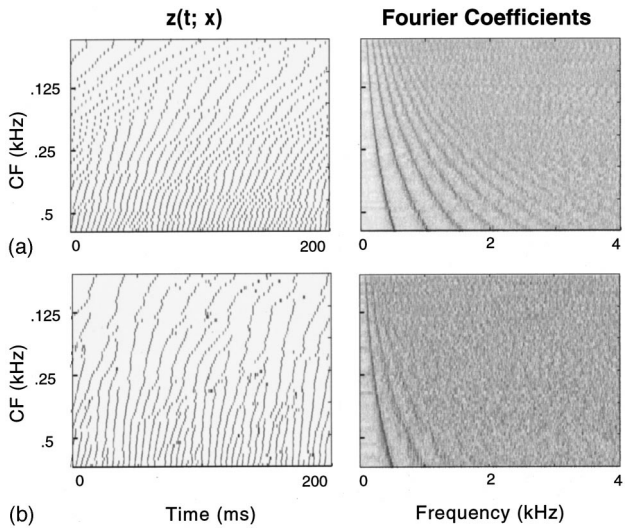

FIG. 6. The effects of spectral resolution on the templates. (Left) The synchronized responses to a broadband noise stimulus [as in Fig. 1(b)]. (Right) The corresponding Fourier series coefficients for all channels (each labeled by its CF along the ordinate). (a) The responses due to the regular model filters (as in Fig. 2). (b) The responses using broader filters (by removing the LIN stage in Fig. 1).

become considerably more jittered due to the increased interference within each channel [Fig. 6(b), left plot]. This in turn smears considerably the Fourier representation of the higher order distortion harmonics [Fig. 6(b), right plot]. Therefore, cochlear frequency selectivity is critical for the formation of the harmonic templates: Sharper filters result in clearer high order harmonic peaks in the templates.

D. Summary

The harmonic templates arise from two basic processing stages: cochlear filtering, followed by a matrix of coincidence detectors. The precise shape of these templates, the clarity of their peaks, and the order of their highest harmonics is influenced by the details of these two operations. The following list summarizes these factors:

(1) Phase-locking of the filter responses is a critical factor in the template formation. All templates are ultimately derived from the fine-time structure of the filter responses. Thus the gradual loss of phase-locking (or synchrony) to higher frequencies (approximately $>2-3$ kHz) is the reason why they are not represented in the templates, and hence play little or no role in the perception of periodicity pitch. In the model, the degree of phase-locking can be simulated by changing the width of the pulse function $p(t)$: the sharper the pulse, the better is the phase-locking to higher frequencies.

(2) Nonlinear transformation of the filter responses is essential in generating the (distortion) harmonics that ultimately form the templates. Half-wave rectification and increasing temporal synchrony are two such transformations. Thus increasing temporal synchrony improves the representation of the higher harmonics.

(3) High spectral resolution improves the representation of the harmonic peaks in the templates. In the model, the lateral inhibitory stage increases the effective tuning of the filters; removing this stage therefore reduces the number and sharpness of the harmonic peaks in the templates.

(4) Phase delays of the traveling wave provide locally phase shifted copies of the responses at each CF. While such phase shifts are typical near the resonance of any bandpass filter, they are especially large in the cochlear filters because of their steep high-frequency roll-off just above the CF. Note that it is important in the model to provide sufficiently dense sampling of the CF axis (number of channels/octave) in order to capture these phase shifts; the sparser the sampling, the weaker are the coincidence peaks in the templates.

(5) The formation of the harmonic templates and their parameters are solely determined by the intrinsic properties of the cochlear filters and coincidences and not of the stimulus. That is, given enough time, the same templates will emerge for any broadband stimulus whether it is noise, harmonic sequences, or impulses.

Note that the combined effect of all these factors give rise to templates (Fig. 3) with features that resemble closely those suggested by some of the algorithmic implementations of the spectral pitch theories [e.g., as in Duifhuis *et al.* (1982); Cohen *et al.* (1995)]. For example, in these implementations, the ideal harmonic templates with their equal amplitude spectral lines (as in Goldstein, 1973b) are modified in two ways: Harmonic peaks are gradually decreased in amplitude and/or increased in width with increasing order. These features arise in our templates due to the various factors discussed above.

Finally, we observe that a close examination of the harmonic templates in Fig. 3 reveals substantially smaller peaks that are interspersed among the harmonic peaks. These peaks are due to sub-harmonic interval correlations. For instance, the 175-Hz fundamental has moderate positive correlation with *approximately* 262 Hz, because these two frequencies are integer multiple harmonics (2nd and 3rd) of *approximately* 87 Hz; other peaks between the first and second harmonics of this template may sometimes be visible at the $\frac{3}{4}, \frac{4}{5}, \cdot$ harmonic ratios. Most other sub-harmonic peaks are much smaller and are rarely evident in our simulations.

The above argument suggests that the templates formed by the coincidence matrix (Fig. 3) are not exactly the ideal harmonic templates hypothesized by classic central pattern matching models (Goldstein, 1973b). Rather, our templates should be more accurately described as equivalent to the ''auto-correlation'' of those ideal templates. However, the computational errors in using our templates in the same manner as an ideal harmonic template are insignificant because the ''nonharmonic'' peaks in our templates are relatively small.

III. PHYSIOLOGICAL CORRELATES OF THE MODEL

We discuss here the biological plausibility of the model and the correspondence between its stages and known physiological responses in the early auditory pathways. Some elements of the model have clear biological underpinnings, while others are speculative. For instance, the frequency analysis, phase shifts around the CF, half-wave rectification, and the phase-locking of the responses are all well known analogs of basilar membrane and hair-cell function.

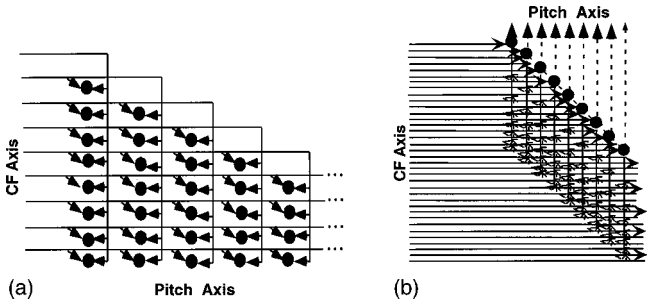

FIG. 7. Biological realizations of the coincidence detectors matrix. (a) The inputs from the auditory channel array are compared pairwise by the network of coincidence detectors. Cells in each column have a common CF input from one side, and a progressively increasing CF input from the other side. The templates emerge along the columns (as illustrated earlier in Fig. 3) when coincidence detectors at harmonic CF distances are strengthened, while others drop out. (b) A different realization where pairwise coincidences are measured and reinforced in the dendrites rather than in separate cells.

More speculative, however, is the anatomy and location of the coincidence matrix, and the identity of its immediate input pathway. Since phase-locking up to relatively high frequencies (at least 2 kHz) is necessary at the input of the matrix, this places it at, or prior to, the inferior colliculus. Furthermore, the synchronized responses at the input of the coincidence matrix are highly reminiscent of the responses of the variety of onset cells in the cochlear nucleus. While these observations suggest certain scenarios as depicted in Fig. 7, the early auditory system is clearly complex and mysterious enough to support many other variant, or even drastically different substrates.

Figure 7 shows two examples of possible ''neural'' realizations of the coincidence matrix. Figure 7(a) is a more literal interpretation of the mathematical model. The matrix consists of tonotopically organized coincidence detectors, where all cells in a column have the same CF, and are also driven by inputs from higher CFs. Thus in the fully formed matrix, each cell ends up driven by a pair of CF inputs: one at its primary CF, and another from a higher, harmonically related CF. In the alternative realization of Fig. 7(b), each cell is driven by its primary CF, but it also has an extensive dendritic tree which spans higher CFs. Initially the dendrites are devoid of synapses. They begin to form during the learning phase at CF locations where the responses correlate well with the primary CF input. In the end, each coincidence cell will be driven by many CF inputs, and hence will appear very broadly tuned. Clearly, a mix of these two scenarios is also possible.

But where is the input pathway to the coincidence matrix? The candidate pathway must be spectrally well resolved and phase-locked as in the auditory nerve. In the cochlear nucleus, many cell types exhibit the appropriate spectrally and temporally sharp responses, especially the onset and primarylike cells in the low CF regions (Rhode, 1994, 1995; Smith and Rhode, 1989; Evans and Zhao, 1998). These cells may project to the coincidence matrix in the limniscus nuclei or the IC. Alternatively, Fig. 7(b) resembles closely the anatomical features of the Octopus cells (the presumed onset-I cells) (Oertel *et al.*, 1990; Palmer *et al.*, 1995), suggesting that they may serve themselves as the coincidence matrix. Unfortunately, most data available at present from various onset cells and other appropriate cell types in the cochlear nucleus are from units with relatively high CFs (>3 kHz), and hence one cannot be certain of their role in periodicity pitch (Palmer *et al.*, 1995; Rhode, 1995; Evans and Zhao, 1998). For instance, the strong dependence of onset cells (especially onset-I) on the phases of the components a complex tone stimulus observed in high CF cells may not occur in low CF cells (Evans and Zhao, 1998).

There are numerous pitch phenomena that are closely related to periodicity pitch, and derived exclusively from binaural stimuli (such as the Huggins pitch[3]). These results suggest that the coincidence detectors may be located at or post binaural convergence nuclei. For instance, it is conceivable that the MSO can serve both its traditional binaural coincidence role (Jeffress, 1948; Shamma *et al.*, 1989), and a monaural coincidence role for the encoding periodicity pitch. Clearly, there is little solid support at present to indicate the existence of such structures in the IC or other central nuclei, and the only definite conclusion that can be made at this time is that much more physiological data are needed to disprove any of these hypotheses.

IV. RESIDUE PITCH

Humans perceive a clear ''residue'' pitch from tone complexes of unresolved components that is equal to the period of the waveform envelope. Unlike periodicity pitch, this percept is sensitive to the phase of the components and is weakest when they are in random phase. It also has different psychoacoustical properties, e.g., bigger **dl**'s, and a different dependence on tone duration (Carlyon, 1998b). This pitch is not related to any harmonicity in the stimulus. There is, therefore, little reason to assume that this pitch is derived from the harmonic templates; instead, it may have a different origin and neural mechanisms, a conclusion also supported psychoacoustically (Carlyon, 1998a). Nevertheless, the need to unify these two pitch percepts in a single mechanism has been a strong motivation for the development of the ''temporal'' models of pitch alluded to earlier.

A. Representing residue pitch in the coincidence matrix

It is possible to show that a simple scheme based on the coincidence matrix is also capable of measuring residue pitch. The basic idea is illustrated in Fig. 8 for a harmonic series stimulus consisting of in-phase high (unresolved) harmonics (10th–110th harmonics of a 70-Hz fundamental). Figure 8(a) shows the synchronized responses evoked by this stimulus (generated by the cochlear model described in Sec. I). A necessary additional ingredient for this scheme to work is a simple monotonic relative increase in response latency from high-to-low CFs at a rate of a few milliseconds per octave [e.g., 6 ms per octave in Fig. 8(b)], similar to that found or postulated in the IC and cortex (Langner and Schreiner, 1988; Greenberg *et al.*, 1998; Hattori and Suga, 1997). This latency-shift effectively delays the response waves in the low CF channels relative to the high CF as shown in Fig. 8(b) causing them to overlap and coincide across some channels. The coincidence matrix indicates the

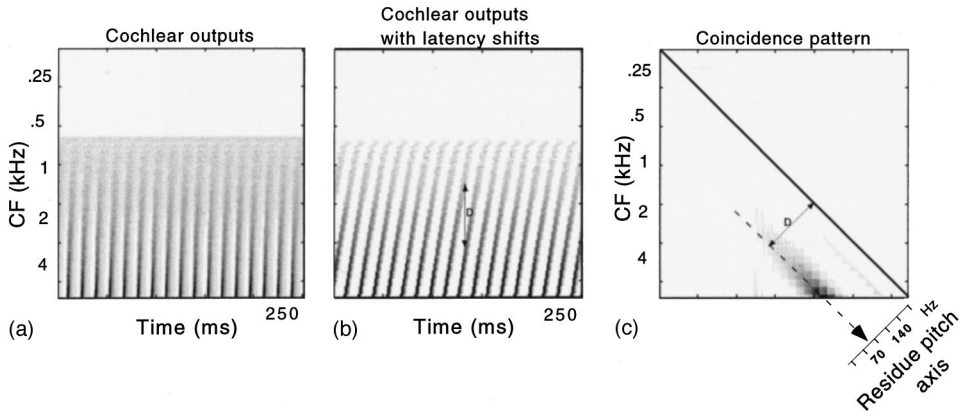

FIG. 8. Measuring residue pitch with the coincidence matrix. (a) The cochlear responses to a stimulus consisting of 10th–110th harmonics of 70 Hz. (b) The responses are systematically delayed by a gradual increase in latency from high to low CF channels. (c) The average output of the coincidence matrix shows a strong line of coincidences parallel to the diagonal, and at a distance (D) that reflects indirectly the period of the stimulus [as indicated in (b)]. Faster rates cause this distance (or line of coincidences) to move gradually closer to the main diagonal as indicated by the pitch axis.

repetition period of the responses in terms of the distance (D) separating the coincident channels in the input array. Thus different periods evoke different coincidence patterns, with faster rates (e.g., 140 Hz) causing coincidences closer to the center diagonal as illustrated in Fig. 8(c).

If the stimulus contains both resolved and unresolved harmonics, both will be simultaneously represented in the coincidence matrix outputs. For example, Fig. 9(a) (left panel) shows the outputs evoked by a stimulus composed of the 1st–31st harmonics of a 250-Hz fundamental. Two distinct coincidence patterns emerge. The first is the diagonal stripes in the high (unresolved) harmonics region (highlighted within the dashed circle). The other is the harmonically spaced ''patches'' in the low (resolved) harmonic region (outside of the dashed circle). Note that the borders of these two regions depend on the fundamental frequency of the stimulus. For instance, for a fundamental frequency of 70 Hz, the coincidence matrix output is dominated by the ''residue pitch'' diagonal patterns (center panel). The opposite is true for a 500-Hz fundamental stimulus (right panel) where only the pattern of resolved harmonic patches is evident.

Many of the well-studied properties of these two types of pitch percepts can be readily seen in the coincidence patterns. For instance, consider the sensitivity of residue pitch to the phases of the unresolved harmonic components. If the phases are completely randomized, the synchronization of the cochlear responses to unresolved harmonics is severely disrupted, leading to the loss of the striped patterns in the output of the coincidence matrix as illustrated in the highlighted region of Fig. 9(b) (left panel) and the entire pattern in Fig. 9(b) (center panel). These figures also illustrate that the patterns in the resolved harmonics region are, as expected, insensitive to phase randomization [Fig. 9(b), left and right panels].

B. Amplitude modulated and rippled noise

Figure 10 illustrates how the coincidence matrix output represents the pitch percepts of two different broadband noise stimuli. In Fig. 10(a) the stimulus is an amplitude-modulated white noise with a modulation rate of 100 Hz (waveform below left panel). The cochlear responses are reminiscent of those due to unresolved harmonics (left panel), and so is the striped output pattern of the coincidence matrix (right panel). The second stimulus is the ''iterated ripple noise'' [Fig. 10(b)]. It is generated by adding (or sub-

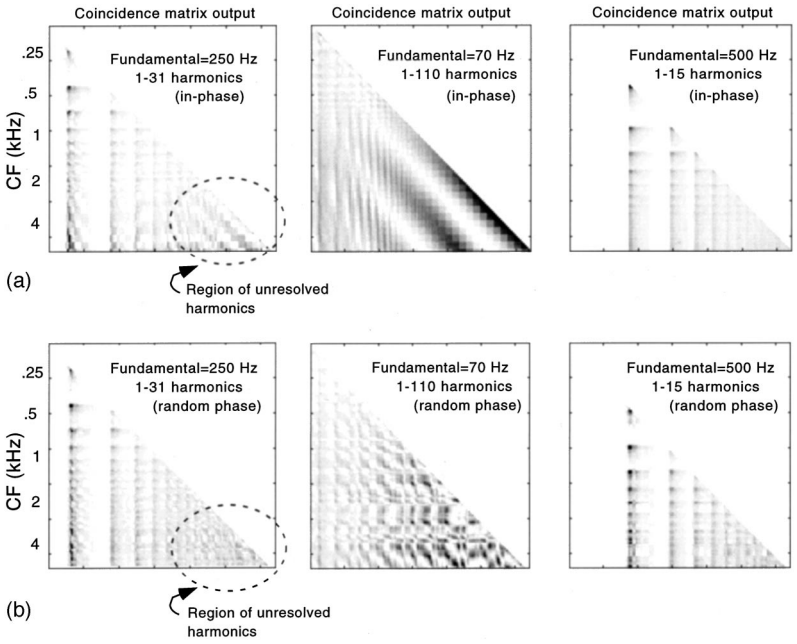

FIG. 9. The representation of resolved and unresolved harmonics in the coincidence matrix outputs. (a) The coincidence matrix outputs for in-phase harmonic series of different fundamental frequencies. (Left panel) A partially resolved harmonic series of 1st–32nd harmonics of 250 Hz. The high-order unresolved harmonics give rise to a striped pattern which is highlighted by the dashed circle in the figure. (Center panel) A mostly unresolved harmonic series of 1st–110th harmonics of 70 Hz. The striped pattern dominates the output. (Right panel) A mostly resolved harmonic series of 1st–15th harmonics of 500 Hz. (b) The coincidence matrix outputs for the same harmonic series stimuli as above, but with randomized phases. The striped pattern due to the unresolved components disappears in left and center panels.

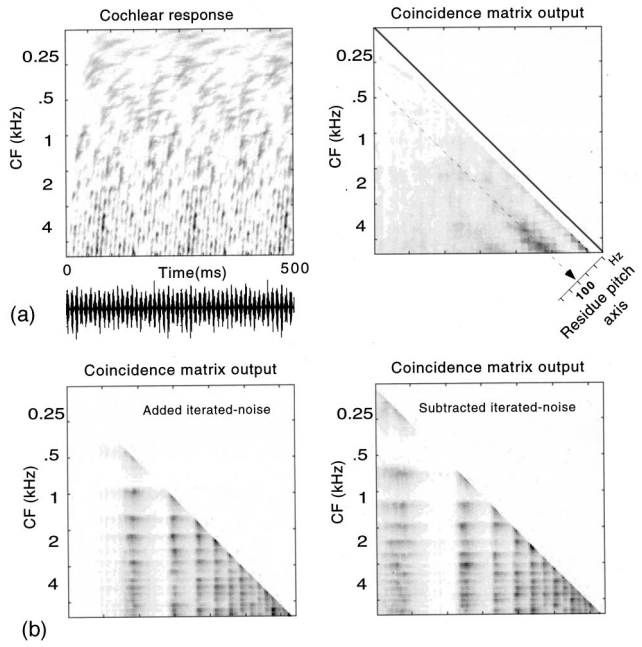

FIG. 10. The representation of pitch due to broadband noise stimuli. (a) Residue pitch evoked by amplitude-modulated broadband noise. (Left panel) Cochlear responses to a 100-Hz amplitude modulated noise (waveform shown below the panel). The responses in the high-CF regions are reminiscent of the responses to unresolved harmonics in Fig. 8. (Right panel) The coincidence matrix outputs indicating the location (at 100 Hz) of the coincidence peak along the same residue pitch axis as in Fig. 8. (b) The coincidence matrix outputs for iterated ripple noise constructed by delaying the noise (2 ms) and then adding (left panel) or subtracting (right panel) the noise to itself. The patterns resemble those of resolved components.

tracting) a delayed version of a white noise to itself several times (Yost and Hill, 1979). The stimulus in Fig. 10(b) is delayed by 2 ms, and added (or subtracted) to itself 16 times. The spectrum of such a stimulus has equally spaced peaks that are $1/d$ Hz apart (Yost and Hill, 1979). The coincidence matrix output exhibits a patchy appearance, and hence we interpret it as evoking *only* a percept of periodicity pitch (and not of residue pitch). Note that the added-noise stimulus (left panel) evokes a percept and a coincidence pattern similar to that of a resolved harmonic series of fundamental $=500$ Hz [see also Fig. 9(c)]. By comparison, the subtracted-noise stimulus (right panel) resembles the *inharmonic* series (250,750,1250,… Hz) and evokes a correspondingly different coincidence pattern.

C. Computing pitch values

The coincidence matrix can potentially be used to compute both ''periodicity'' and ''residue'' pitches depending on how the output is integrated. For periodicity pitch, the output is summed using the learned harmonic templates shown in Fig. 3 [or more graphically as in Fig. 7(b)]. For residue pitch, the output is summed along the diagonals illustrated by the dashed line in Figs. 8(c) and 10(a). Clearly, each of these two types of outputs contributes optimally only to one of the summation methods. Thus the striped pattern contributes little information if summed according to the harmonic templates. Similarly, the harmonic intervals between the patches are irrelevant along the diagonals.

D. Summary

The illustrations in Figs. 8–10 suggest that the coincidence matrix may serve as a common computational mechanism for both periodicity and residue pitch provided that two different summation strategies are employed: the harmonic templates for periodicity pitch and the diagonals for residue pitch. It is also possible that these two computations are segregated in the auditory system into two independent coincidence networks, where the representation of each pitch percept is independently optimized and only one summation strategy is used. The results from these two coincidence networks could subsequently be combined and registered relative to each other (Carlyon, 1998a).

V. DISCUSSION

We have described a model for how harmonic templates might arise during early development of the auditory system. The model demonstrates that the templates are a natural consequence of basic properties of processing in the early stages of the auditory system. Most important among these properties are cochlear filtering, phase-locked representation of its outputs, enhanced temporal synchrony, and, finally, coincidences across the channel array. We have discussed the contributions of each of these properties to the clarity of the template peaks and the highest harmonic order represented.

An important conclusion from this model is that the harmonic templates are robust and reflect fundamental features of peripheral auditory function. Thus for the model to work at all, we must have cochlear frequency analysis; we must have rapid traveling wave delays near the wave's resonance; and we must have phase-locking and half-wave rectification on the auditory nerve. Beyond these fundamental features, all other details, such as enhanced temporal synchrony and spectral sharpness, are helpful in improving the templates in a graded fashion.

Another important conclusion is that template formation is largely independent of the stimulus as long as there is energy available at all frequencies (<3 kHz) over a period of time. That is, harmonic templates will appear if we had used harmonic sounds, impulses, or any other broadband stimulus provided that all frequencies are represented over the ensemble. However, even if the stimulus energy is not well balanced due, for instance, to partial threshold elevation or a notch in the audiogram, the templates will still arise, but with reduced contributions from these frequencies. For example, if the channel at CF$=$400 Hz is removed at the outset (e.g., due to a localized hair cell death at that location), then the model predicts that the 400-Hz template will not be learned, and that this pitch will not be heard from a complex of higher order harmonics (e.g., 800, 1200, 1600 Hz). All other templates will form, but with contribution from the 400 Hz missing. For instance, the 200-Hz template will have all its peaks intact except for the 400 Hz. Note that this prediction is contrary to that obtained from a ''temporal'' model such as the correlogram (Slaney and Lyon, 1993), where the perceptual contribution to the 400-Hz pitch comes from all CF channels regardless of what is happening at the CF$=$400 Hz channel.

Finally, the model suggests a simple answer to the question of why harmonic templates postulated in psychoacoustic studies have a prominent fundamental when natural harmonic sounds (e.g., speech) often have little or no energy at the fundamental? Equivalently, why does a partial set of upper harmonics evoke a pitch at the fundamental and not at any other arbitrary frequency, thus implying that the learned templates must be linked to the fundamental? The answer is that harmonic templates are formed from exposure to broadband noise, clicks, and other stimuli where all frequencies are available, and not simply from examples of harmonic sounds such as voices which may not have the fundamental.

A. Where to search for physiological evidence

What physiological or anatomical evidence should we look for to confirm the presence of the harmonic templates? Two sets of data are needed to shed light on the model. The first concerns the inputs to the coincidence matrix, and the second deals with the coincidence cells themselves. The input pathway must be sharply tuned in its *synchronous responses*. This is an important consideration which is often ignored when reporting on the tuning or iso-intensity response curves of these cochlear units. To establish the relevance of any cells for the encoding of periodicity pitch, it is essential that the units studied receive phase-locked auditory-nerve inputs, and hence must have low CFs (<3 kHz). It is also best if their tuning properties are measured with reference to their phase locked, and not their average rate, inputs. For example, an onset cell may appear very broadly tuned due to its relatively high threshold and very limited dynamic range (Rhode, 1995; Evans and Zhao, 1998). However, the unit may be sharply tuned if one considers how well it is synchronized to one of several closely spaced stimulus components (Greenberg *et al.*, 1986). Alternatively, onset cells may constitute the coincidence matrix themselves, and hence receive input from several CFs (e.g., a pair), and appear broadly tuned.

Coincidence detectors, wherever they may reside, should exhibit distinctive response patterns to harmonic series stimuli such as click trains. For instance, coincidence cells as in Fig. 7(b) should be selective to the rate of a click train (tuned to the fundamental of the cell's template); they must also be insensitive to the phase of the harmonics in the stimulus; and finally, they should be broadly tuned, or at least broadly facilitated (Palmer *et al.*, 1995). The response patterns are different if the coincidence cells have pairwise inputs as in Fig. 7(a). The cells should be doubly tuned or facilitated, and must exhibit predictable tunings to multiple click rates in a phase-insensitive manner. None of these response properties have been reliably demonstrated in the IC or lower auditory nuclei, and it remains to be seen if more controlled recordings in the low CF regions can shed light on these questions.

B. The principle of coincidence detection

In biologically inspired models of various auditory tasks, it has been common to postulate neural delays (prior to coincidence detection) so as to affect various correlation operations. These delays are explicit in some algorithms, e.g., in the binaural processing of interaural time delays (Jeffress, 1948; Colburn and Durlach, 1978), or in computing the correlograms for pitch (Slaney and Lyon, 1993). In other algorithms, these delays are implicit only within purely temporal operations that must use them, e.g., in the ALSR and dominant frequency algorithms for spectral shape extraction from auditory-nerve responses (Young and Sachs, 1979; Lyon and Shamma, 1996), or in the use of intrinsic oscillations or firing intervals of variable rates for pitch estimation (Langner, 1992; Hewitt and Meddis, 1994; Winter *et al.*, 1999). As mentioned earlier, the need for neural delays stems almost entirely from the need to make interval measurements on single channels independent of other channels.

In previous reports, we have demonstrated that simple coincidence measurements of responses *across* the auditory channels can extract the same kinds of information robustly, without need for functional neural delays. Thus lateral inhibition across the outputs of the auditory-nerve fiber array (which is essentially a form of coincidence detection) can extract a highly resolved spectrum of a broadband complex stimulus over wide stimulus levels (Shamma, 1985a, b). Similarly, a coincidence matrix identical to the one discussed here (Fig. 1) for auditory channels from the two ears [the stereausis network (Shamma *et al.*, 1989)] potentially can explain binaural phenomena accounted for by traditional cross-correlation models. The algorithm described in this paper repeats the same theme discussed above. That is, coincidences across the fiber array carry sufficient information to generate the harmonic templates, and hence obviate the need for the neural delay lines invoked in many of the current pitch models.

To summarize, the need to invoke neural delay lines stems from a common view of auditory processing as primarily temporal in the sense defined earlier. This view may be partially a consequence of the experimental difficulty of measuring the distribution of auditory responses across the tonotopic axis, and hence of appreciating the richness and subtlety of the *spatio*temporal cues created by the cochlea. Decoding such cues often requires simple coincidence detection to measure time differences between channels rather than absolute time intervals within a channel. Temporal models, instead, essentially recreate cochlealike frequency analysis centrally by postulating additional ordered delay lines, correlators, and narrowly tuned filters with fine (microsecond) accuracy to detect and measure response periodicities and stimulus parameters.

C. Using the model to compute pitch

This paper has primarily dealt with the question of how the harmonic templates associated with periodicity pitch emerge as a natural consequence of cochlear function and subsequent coincidence networks. We have also discussed in passing the representation of residue pitch in the coincidence network (Sec. IV), and hence the potential role of coincidence as a unifying mechanism underlying both types of pitch percepts.

It is possible to formulate a complete and detailed computational model of pitch perception based on these mechanisms. The model would conceptually consist of two parts:

(1) computation of periodicity pitch based on the learned harmonic templates; and (2) a computation of residue pitch based on the diagonal patterns in the coincidence matrix outputs. As eluded to earlier, the periodicity pitch portion of this model has existed for decades in the literature as template matching algorithms of various forms (Duifhuis *et al.*, 1982; Cohen *et al.*, 1995). For instance, Cohen and Grossberg (1995) presented a detailed account of the psychoacoustical data that could be accounted for by the harmonic template match model, including the ambiguous pitches and phase insensitivity. Undoubtedly, further refinements on this part of the model can be made in the future. For example, implementing the matching operation between the input spectra and the harmonic templates as essentially a simple cross correlation often produces multiple additional estimates at the octaves and other intervals (Duifhuis *et al.*, 1982; Cohen *et al.*, 1995; Goldstein, 1973a). While subjects often report these pitches under experimental scrutiny, pitch percepts in casual listening conditions are usually more unitary (or less analytical), a property that may reflect subsequent integrative processing at higher auditory centers.

The residue pitch portion of our model (Sec. IV) complements the template matching algorithm. It accounts for the phase sensitivity and other properties of pitch percepts that *do not* involve any resolved harmonics such as the pitch of high order harmonics and amplitude modulated noise. However, two important issues remain to be addressed in the future. The first is whether the coincidence algorithm for residue pitch can withstand a critical quantitative scrutiny of its properties, similar to that done earlier for the template matching algorithm (Cohen *et al.*, 1995; Goldstein, 1973a). The second issue concerns the exact nature of the integration of the two pitch percepts. The relevance of this issue stems from the fact that natural stimuli commonly contain simultaneous cues for both pitch percepts (e.g., resolved and unresolved harmonics). Consequently, the final unitary pitch percept must be derived from both available cues.

VI. SUMMARY AND CONCLUSIONS

We have presented a biologically plausible model for forming harmonic templates in the early stages of the auditory system *with broadband noise stimulation, and without need for neural delay lines and other temporal structures*. The model consists of two key operations: a cochlear filtering stage followed by coincidence detection. The cochlear stage provides responses analogous to those seen in the auditory nerve and cochlear nucleus. The second stage is a matrix of coincidence detectors that compute the long-term average of pairwise instantaneous correlation (or products) between responses from all CFs across the channels. Model simulations show that for any broadband stimulus, high coincidences occur between cochlear channels that are exactly harmonic intervals apart. Accumulating coincidences over time results in the formation of harmonic templates for all fundamental frequencies in the phase-locking frequency range. The model explains the critical role played by such important factors in cochlear function as the nonlinear transformations following the filtering stage, the rapid phase-shifts of the traveling wave near its resonance, and the spectral resolution of the cochlear filters. More specifically, the following items summarize the major findings of the model:

(1) Phase-locking of the filter responses is a critical factor in the template formation. All templates are ultimately derived from the fine-time structure of the filter responses. Thus the gradual loss of phase locking (or synchrony) to higher frequencies (approximately $>2-3$ kHz) is partially the reason why they are not represented in the templates, and hence play little or no role in the perception of periodicity pitch.

(2) Nonlinear transformation of the filter responses is essential in generating the ''distortion'' harmonics that ultimately form the templates. Half-wave rectification and increasing temporal synchrony are two such transformations.

(3) High spectral resolution improves the representation of the harmonic peaks in the templates. Broadening the analysis filters smears the representation of the higher order harmonics considerably.

(4) Phase delays of the traveling wave provide locally phase shifted copies of the responses at each CF, which in turn insures there is always a pair of channels at harmonic CFs that can be highly correlated regardless of the phase of the stimulus harmonics.

(5) The exact nature of the sound stimulus is immaterial for the formation of the harmonic templates and their parameters. Instead they are solely determined by the intrinsic properties of the cochlear filters and coincidences. That is, given enough time, the same templates will emerge for any broadband stimulus whether it is noise, harmonic sequences, or impulses.

ACKNOWLEDGMENTS

We are grateful to Dr. Steve Greenberg and Dr. Ray Meddis for extensive reviews and advice. This work was supported in part by a grant from the Office of Naval Research under the ODDR&E MURI97 Program to the Center for Auditory and Acoustic Research, and the National Science Foundation under the Learning and Intelligent Systems Initiative Grant No. CMS9720334.

[1]The use of harmonic templates in Wightman's model is somewhat less obvious than in others. Wightman's model utilizes a linear frequency axis to represent the spectrum, and then performs a cepstral analysis (Oppenheim and Schafer, 1976) to compute the inter-harmonic distances (and hence the pitch). Therefore, the harmonic templates used are (implicitly) the Fourier basis functions used in the cepstral analysis.

[2]The distinction we make here between ''spectral'' and ''temporal'' pitch theories is not universally accepted. Our use of the terms here is rather specific. We define ''spectral pitch theories'' as those that take as their starting point a spectral pattern from which the pitch is computed *regardless* of the nature of the preceding cues and mechanisms used to extract this spectral pattern. For example, it is immaterial to our definition whether the spectrum is extracted from the average auditory-nerve firing rates (Sachs and Young, 1979), time-intervals (Seneff, 1988; Ghitza, 1988), or other cues (Young and Sachs, 1979; Shamma, 1985b). ''Temporal pitch theories'' do not make use of a spectral pattern; instead, they utilize temporal response features from each auditory-nerve fiber independently of other channels. Therefore, unlike in ''spectral'' theories, tonotopic order plays no role in ''temporal'' theories; i.e., shuffling around the auditory-nerve fiber array destroys the spectral pattern, but makes no difference to the temporal patterns on each channel.

[3]I am indebted to Dr. Steve Greenberg for bringing to my attention an interesting link between the rapid phase shifts of the cochlear traveling

wave (discussed earlier in Sec. II and Fig. 5) and the *Huggins pitch* (Cramer and Huggins, 1958). The *Huggins pitch* is a ''tonal'' percept created by local phase shifts in the spectrum of otherwise uncorrelated noise presented to the two ears. The saliency of the percept is crucially dependent on the spectral interval over which the phase shifts occur—it is best when they occur over an interval of 3–6 percent of the center frequency of the relevant filter. This estimate may reflect the region over which the rapid phase shifts occur, and indirectly measures the spectral resolution expected from such percepts (i.e., the smallest distance between two locally imposed phase shifts). Bilsen (1977) has demonstrated that multiple phase shifts create a multi-tonal percept which, if harmonically related, evoke a missing fundamental pitch percept. As Bilsen points out (Bilsen, 1977), these results are consistent with a template matching pitch algorithm in which the input spectrum is derived centrally from the binaural inputs which displays peaks corresponding to the tonal Huggins pitch percepts. How this spectrum is derived is beyond the scope of this paper, but traditional binaural coincidence algorithms can readily perform this task.

Bilsen, F. (**1977**). ''Pitch of noise signals: Evidence for a central spectrum,'' J. Acoust. Soc. Am. **61**, 150–161.

Cariani, P., and Delgutte, B. (**1996a**). ''Neural correlates of the pitch of complex tones. i: Pitch and pitch salience,'' J. Neurophysiol. **76**, 1698–1716.

Cariani, P., and Delgutte, B. (**1996b**). ''Neural correlates of the pitch of complex tones. ii: Pitch shift, pitch ambiguity, phase invariance, pitch circularity, rate pitch, and the dominance region for pitch,'' J. Neurophysiol. **76**, 1717–1734.

Carlyon, R. (**1998a**). ''Comments on a unitary model of pitch perception,'' J. Acoust. Soc. Am. **104**, 1118–1121.

Carlyon, R. (**1998b**). ''The effects of resolvability on the encoding of fundamental frequency by the auditory system,'' in *Psychophysical and Physiological Advances in Hearing*, edited by A. R. Palmer, A. Rees, A. Q. Summerfield, and R. Meddis (Whurr, London).

Clarkson, M., and Rogers, E. (**1995**). ''Infants require low-frequency energy to hear the pitch of the missing fundamental,'' J. Acoust. Soc. Am. **98**, 148–154.

Cohen, M., Grossberg, S., and Wyse, L. (**1995**). ''A spectral network model of pitch perception,'' J. Acoust. Soc. Am. **98**, 862–879.

Colburn, S., and Durlach, N. (**1978**). ''Models of binaural interactions,'' in *Handbook of Perception*, edited by E. Carterette and M. Friedman (Academic, New York), Vol. IV.

Cramer, E., and Huggins, W. (**1958**). ''Reaction of pitch through binaural interactions,'' J. Acoust. Soc. Am. **30**, 413–417.

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

de Cheveigne, A. (**1998**). ''Cancellation model of pitch perception,'' J. Acoust. Soc. Am. **103**, 1261–1271.

de Cheveigne, A., McAdams, S., and Marin, C. (**1995**). ''Concurrent vowel identification. ii: Effect of phase, harmonicity, and task,'' J. Acoust. Soc. Am. **101**, 2848–2856.

Duifhuis, H., Willems, L., and Sluyter, R. (**1982**). ''Measurement of pitch in speech: An implementation of Goldstein's theory of pitch perception,'' J. Acoust. Soc. Am. **71**, 1568–1580.

Evans, E. (**1978**). ''Place and time coding in the peripheral auditory system: Some physiological pros and cons,'' Audiology **17**, 369–420.

Evans, E., and Zhao, W. (**1998**). ''Periodicity coding of the fundamental frequency of harmonic complexes. Physiological and pharmalogical study of onset units in the ventral cochlear nucleus,'' in *Psychophysical and Physiological Advances in Hearing. Proceedings of the 11th International Symposium on Hearing*, edited by A. R. Palmer, A. Ress, A. Q. Summerfield, and R. Meddis (Whurr Publishers, London), pp. 186–194.

Ghitza, O. (**1988**). ''Temporal non-place information in the auditory-nerve firing patterns as a front-end for speech recognition in a noisy environment,'' J. Phonetics **16**, 109–204.

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

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

Greenberg, S., Geisler, C., and Deng, L. (**1986**). ''Frequency selectivity of single cochlear nerve fibers based on the temporal response patterns of two-tone signals,'' J. Acoust. Soc. Am. **79**, 1010–1019.

Greenberg, S., Poeppel, D., and Roberts, T. (**1998**). ''A space-time theory of pitch and timbre based on cortical expansion of the cochlear travelling-wave delay,'' in *Psychophysical and Physiological Advances in Hearing. Proceedings of the 11th International Symposium on Hearing*, edited by A. R. Palmer, A. Ress, A. Q. Summerfield, and R. Meddis (Whurr Publishers, London).

Hattori, T., and Suga, N. (**1997**). ''The inferior colliculus of the mustached bat has the frequency-vs-latency coordinates,'' J. Comp. Physiol. A **180**, 271–284.

Hewitt, J., and Meddis, R. (**1994**). ''A computer model of amplitude-modulation sensitivity of single units in the inferior colliculis,'' J. Acoust. Soc. Am. **95**, 2145–2159.

Jeffress, A. (**1948**). ''A place theory of sound localization,'' J. Comp. Physiol. Psychol. **61**, 468–486.

Langner, G. (**1992**). ''Periodicity coding in the auditory system,'' Hear. Res. **6**, 115–142.

Langner, G., and Schreiner, C. (**1988**). ''Periodicity coding in the inferior colliculus of the cat,'' J. Neurophysiol. **60**, 1805–1822.

Lickleider, J. (**1951**). ''A duplex theory of pitch perception,'' Experientia **7**, 128–133.

Lyon, R., and Shamma, S. (**1996**). ''Auditory representation of timbre and pitch,'' in *Auditory Computations*, edited by H. Hawkins, E. T. McMullen, A. Popper, and R. Fay (Springer-Verlag, Berlin), pp. 221–270.

Meddis, R., and Hewitt, J. (**1991**). ''Virtual pitch and phase sensitivity of a computer model of the auditory periphery. i: Pitch identification,'' J. Acoust. Soc. Am. **89**, 2866–2882.

Montgomery, C., and Clarkson, M. (**1997**). ''Infants' pitch perception: Masking by low- and high-frequency noises,'' J. Acoust. Soc. Am. **102**, 3665–3672.

Moore, B. (**1989**). *An Introduction of the Psychology of Hearing*, 3rd ed. (Academic, London).

Moore, B. C. J. (**1986**). *Frequency Selectivity in Hearing* (Academic, London), Chap. 5.

Oertel, D., Wu, S., and Dizack, C. (**1990**). ''Morphology and physiology of cells in slice preparation of posterovental cochlear nucleus of mice,'' J. Comp. Neurol. **295**, 136–154.

Oppenheim, A., and Schafer, R. (**1976**). *Digital Signal Processing* (Prentice-Hall, New Jersey).

Palmer, A., Winter, I., Jiang, D., and James, N. (**1995**). ''Across frequency integration by neurons in the ventral cochlear nucleus,'' in *Advances in Hearing Research*, edited by J. Manley, G. Klump, C. Kopple, H. Fastl, and H. Oeckinghaus (World Scientific, Singapore).

Patterson, R., and Holdsworth, J. (**1991**). ''A functional model of neural activity patterns and auditory images,'' in *Advances in Speech, Hearing and Language Processing*, edited by W. A. Ainsworth (JAI Press, London), Vol. 3.

Plomp, R. (**1976**). *Aspects of Tone Sensation* (Academic, New York).

Rhode, W. (**1994**). ''Lateral suppression and inhibition in the cochlear nucleus of the cat,'' J. Neurophysiol. **71**, 493–519.

Rhode, W. (**1995**). ''Interspike intervals as a correlate of periodicity pitch in cat cochlear nucleus,'' J. Acoust. Soc. Am. **97**, 2414–2429.

Ruggero, M. (**1973**). ''Response to noise in auditory nerve fibers in squirrel monkey,'' J. Neurophysiol. **36**, 569–587.

Sachs, M. B., and Young, E. D. (**1979**). ''Encoding of steady state vowels in the auditory-nerve: Representation in terms of discharge rate,'' J. Acoust. Soc. Am. **66**, 470–479.

Schouten, J. (**1940**). ''The residue and the mechanism of hearing,'' Proc. K. Ned. Akad. Wet. **43**, 991–999.

Schreiner, C., and Langner, G. (**1988**). ''Periodicity coding in the inferior colliculus of the cat. ii. topographical organization,'' J. Neurophysiol. **60**, 1823–1840.

Schreiner, C., and Urbas, J. (**1988**). ''Representation of amplitude modulation in the auditory cortex of the cat. i: The anterior field,'' Hear. Res. **21**, 227–241.

Schwartz, D., and Tomlinson, R. (**1990**). ''Spectral response patterns of auditory cortex neurons to harmonic complex tones in alert monkey (*macaca mulatta*),'' J. Neurophysiol. **64**, 282–299.

Seneff, S. (**1988**). ''A joint synchrony/mean-rate model of auditory processing,'' J. Phonetics **85**, 55–76.

Shamma, S. (**1985a**). ''Speech processing in the auditory system: I. representation of speech sounds in the responses of the auditory nerve,'' J. Acoust. Soc. Am. **78**, 1612–1621.

Shamma, S. (**1985b**). ''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**, 1622–1632.

Shamma, S., Chadwick, R., Wilbur, J., Morrish, K., and Rinzel, J. (**1986**). ''A biophysical model of cochlear processing: Intensity dependence of pure tone responses,'' J. Acoust. Soc. Am. **80**, 133–145.

Shamma, S., and Morrish, K. (**1986**). ''Synchrony suppression in complex stimulus responses of a biophysical model of the cochlea,'' J. Acoust. Soc. Am. **81**, 1486–1498.

Shamma, S., Shen, N., and Gopalaswamy, P. (**1989**). ''Stereausis: Binaural processing without neural delays,'' J. Acoust. Soc. Am. **86**, 989–1006.

Slaney, M., and Lyon, R. (**1993**). ''On the importance of time—A temporal representation of sound,'' in *Visual Representations of Speech Signals*, edited by M. Cooke, S. Beet, and M. Crawford (Wiley, New York).

Smith, P., and Rhode, W. (**1989**). ''Structural and functional properties distinguish two types of multipolar cells in the cat ventral cochlear nucleus,'' J. Acoust. Soc. Am. **282**, 595–616.

Summerfield, A., and Assmann, P. (**1990**). ''Modelling the perception of concurrent vowels: Vowels with different fundamental frequencies,'' J. Acoust. Soc. Am. **88**, 680–697.

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

Wang, K., and Shamma, S. A. (**1994**). ''Self-normalization and noise-robustness in early auditory representations,'' IEEE Trans. Speech Audio Process. **2**, 421–435.

Wightman, F. (**1973**). ''A pattern transformation model of pitch,'' J. Acoust. Soc. Am. **54**, 397–406.

Winter, I., Wiegrebe, L., and Patterson, R. (**1999**). ''Encoding iterated ripple noise and harmonic complexes in the ventral cochlear nucleus,'' in *Abs. 552, Assn. Res. Otol. Annual Meeting*.

Yang, X., Wang, K., and Shamma, S. A. (**1992**). ''Auditory representations of acoustic signals,'' IEEE Trans. Inf. Theory, Special Isssue on Wavelet Transforms and Multiresolution Signal Analysis **38**, 824–839.

Yost, W., and Hill, R. (**1979**). ''Model of the pitch and pitch strength of ripple-noise,'' J. Acoust. Soc. Am. **66**, 400–410.

Young, E., and Sachs, M. (**1979**). ''Representation of steady-state vowels in the temporal aspects of the discharge patterns of populations of auditory-nerve fibers,'' J. Acoust. Soc. Am. **66**, 1381–1403.