The case of the missing delay lines: Synthetic delays obtained by cross-channel phase interaction

Alain de Cheveigné and Daniel Pressnitzer
Equipe Audition, FRE 2929, CNRS, Université Paris 5, ENS, 29 Rue d'Ulm, F-75230 Paris cedex 05, France

(Received 6 February 2005; revised 18 March 2006; accepted 22 March 2006)

Temporal models of pitch and harmonic segregation call for delays of up to 30 ms to cover the full range of existence of musical pitch. To date there is little anatomical or physiological evidence for delays that long. We propose a mechanism by which delays may be synthesized from cross-channel phase interaction. Phases of adjacent cochlear filter channels are shifted by an amount proportional to frequency and then combined as a weighted sum to approximate a delay. Synthetic delays may be used by pitch perception models such as autocorrelation, segregation models such as harmonic cancellation, and binaural processing models to explain sensitivity to large interaural delays. The maximum duration of synthetic delays is limited by the duration of the impulse responses of cochlear filters, itself inversely proportional to cochlear filter bandwidth. Maximum delay is thus frequency dependent. This may explain the fact, puzzling for temporal pitch models such as autocorrelation, that pitch is more salient and easy to discriminate for complex tones that contain resolved partials. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2195291]

PACS number(s): 43.66.Ba, 43.64.Bt, 43.66.Hg [AK] Pages: 3908–3918

I. INTRODUCTION

The autocorrelation (AC) model is a popular account for pitch perception (for a review of pitch models see de Cheveigné, 2005). Initially proposed by Licklider (1951) and later refined by Meddis and Hewitt (1991a, b) and others, the AC model accounts well for a wide range of pitch phenomena and is consistent with the electrophysiological recordings of Cariani and Delgutte (1996a, b). It has however at least two weaknesses. First, in the implementation proposed by Licklider it requires arrays of internal *delays* of up to about 30 ms to account for the full range of musical pitch (Pressnitzer *et al.*, 2001a). Conclusive anatomical and physiological evidence for such delays has not been found to date. Second, it fails to predict the greater pitch salience and discrimination accuracy of stimuli with resolved harmonics, as compared to stimuli that contain only unresolved harmonics (Houtsma and Smurzynski, 1990; Shackleton and Carlyon, 1994; Carlyon and Shackleton, 1994; Oxenham *et al.*, 2004), although this question is still in debate (Meddis and O'Mard, 1997; Carlyon, 1998).[1]

Another popular account of pitch is based on the concept of pattern matching (de Boer, 1956; Goldstein, 1973; Wightman, 1973; Terhardt, 1974): periodicity pitch (also called low, residue, or virtual pitch) is derived from the pattern formed by the frequencies of individual partials. For that, these partials must be *resolved* by peripheral filtering, and thus pattern matching cannot account for the pitch of complexes with only unresolved partials. The pitch evoked by those stimuli is weak but nevertheless musical (Pressnitzer *et al.*, 2001a; Kaernbach and Bering, 2001). To account for pitch over a full range of stimulus conditions, pattern matching must be associated with another mechanism such as AC. This "dual mechanism" hypothesis (Shackleton and Carlyon, 1994) is less parsimonious than a "unitary model" (e.g., Meddis and O'Mard, 1997), but has the appeal that it accounts for a difference in pitch salience between resolved and unresolved complexes.

In addition to pitch, internal delays are invoked for other aspects of auditory processing. The cancellation model of harmonic sound segregation uses a delay equal to the period of an unwanted sound (de Cheveigné, 1993, 1997, 1999). Delays of about 10 ms are needed to account for typical concurrent vowel identification results. Stimuli with interaural delays of up to 10 ms (Mossop and Culling, 1999) or 20 ms (Blodgett *et al.*, 1956) can be discriminated on the basis of laterality. To explain this using a crosscorrelation display model (e.g., Stern and Shear, 1996), this display would need to extend to delays of a similar size. There are however other accounts of laterality that involve shorter delays (Saberi *et al.*, 2001). Other pitch models in addition to autocorrelation require delays, for example the STI (strobed temporal integration) model of Patterson *et al.* (1995), or the cancellation model of de Cheveigné (1998). Long delays are also required by models that attempt to account for the accuracy of pure tone pitch discrimination on the basis of intervals that extend over multiples of the period (e.g., de Cheveigné, 1989, 2000; Slaney, 1990). Delay is an ingredient of many auditory models, and the lack of evidence for neural delays over the range needed is puzzling.

This paper examines a mechanism by which delays can be synthesized by phase interaction between cochlear channels. Interestingly, this mechanism might also account for the resolvability limits of pitch. It does, however, not absolve from the need of a second mechanism involving long physical delays for the pitch elicited by unresolved partials. The next section presents the idea, Sec. III discusses issues that must be addressed before the idea is considered plausible for auditory modeling, and Secs. IV and V, respectively, review relevant psychophysical and physiological data.

0001-4966/2006/119(6)/3908/11/$22.50

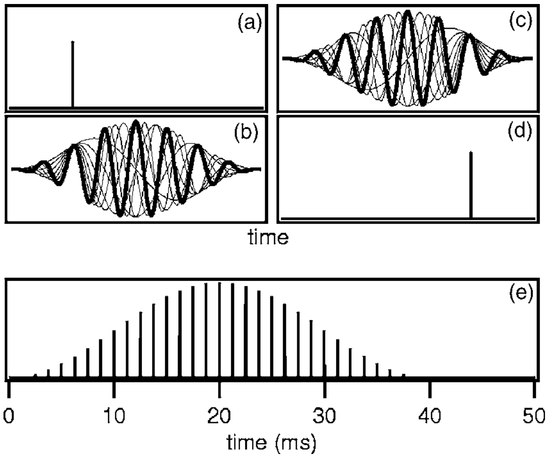

FIG. 1. Illustration of a delay implemented with a short-term Fourier transform. (a) Pulse. (b) Decomposition of the pulse over a basis of windowed sinusoids (first ten terms). (c) Same sinusoids after a phase shift proportional to frequency. (d) Sum of phase-shifted windowed sinusoids. The resulting pulse is delayed with respect to the input pulse. (e) Illustration of the range of pulses that can be synthesized by adding the windowed sinusoids of (b) supposing a 40 ms raised-cosine window.

II. DELAY AS PHASE SHIFT

A delay in the time domain is equivalent to a phase shift proportional to frequency in the Fourier domain. For a delay of D, the phase shift is

$$\phi = \omega D, \tag{1}$$

where $\omega = 2\pi f$. Indeed, if the spectrum of $x(t)$ is $X(\omega)$, the delayed signal is

$$x(t+D) = \int_{-\infty}^{+\infty} [X(\omega)e^{-j\omega D}]e^{-j\omega t}d\omega. \tag{2}$$

Any delay can be obtained in this manner with the Fourier transform. With the *short-term* Fourier transform, in which the signal is decomposed over a basis of windowed sinusoids, a restricted set of delays can be synthesized. Figure 1 illustrates the process: The pulse (a) is analyzed into a sum of windowed sinusoids (b). Sinusoids are phase shifted (c) and then added to produce the delayed pulse (d). Interference between the sinusoids is everywhere destructive except at one time point. By manipulating the phases, the position of this point can be moved within the range covered by the analysis window. Figure 1(e) illustrates the range of pulses that can be synthesized for a raised-cosine window of 40 ms.

The cochlea has been likened to a Fourier transformer (von Helmholz, 1877). Could it be used to support a similar operation? Several differences between auditory peripheral frequency analysis and a short-term Fourier transform need to be considered. Cochlear filter bank parameters (bandwidth, sampling) are not uniform, there are nonlinearities at several stages (cochlear filtering, transduction, and neural processing), and the neural analogue of summation in Eq. (2) is not immediately obvious. We will show that a delaylike operation can nevertheless be approximated using a model of cochlear filtering instead of Fourier analysis.

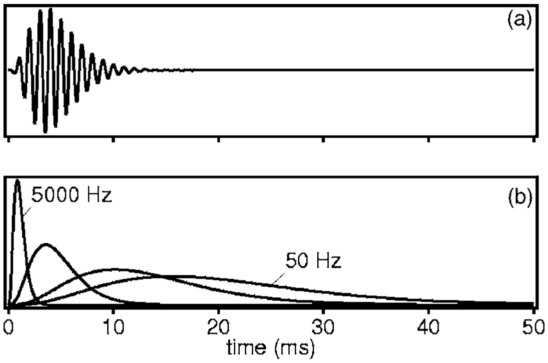

FIG. 2. Top: impulse response of a 1000 Hz gammatone filter. Bottom: envelopes of impulse responses of gammatone filters centered at 50, 200, 1000, and 5000 Hz.

A. Delays produced with a gammatone filterbank

An approximation of cochlear filtering is the gammatone filter (Carney and Yin, 1988). Its impulse response is the product of a sinusoid and an envelope function

$$h(t) = g(t)\cos(2\pi f t + \phi), \tag{3}$$

where f is the center frequency (CF) of the filter, ϕ determines the phase, and

$$g(t) = at^{(n-1)}e^{-2\pi bt} \tag{4}$$

defines the envelope. Parameter n is the order of the filter ($n=4$ in this paper) and b determines the filter bandwidth. For a fourth-order gammatone, b and the equivalent rectangular bandwidth B (ERB, Moore and Glasberg, 1983) are related by $b=1.018B$. Bandwidths of auditory filters follow approximately the formula $B=24.7+0.108f$ where f is the filter center frequency, so that high-frequency channels are wider than low-frequency channels. Figure 2(a) illustrates the impulse response of a gammatone filter centered on 1 kHz, and Fig. 2(b) shows temporal envelopes of impulse responses of selected filters with CFs between 50 and 5000 Hz. Note that low-CF responses peak later and last longer than high-CF responses. In contrast, the basis functions of the short-term Fourier transform illustrated in Fig. 1(b) had a common envelope determined by the windowing function.

We now make two important assumptions. The first is that the phase of each channel of the gammatone filterbank [ϕ in Eq. (3)] may be adjusted arbitrarily. Specifically, we need to set the phase of each channel k to $\phi_k = -2\pi D f_k$, where f_k is the CF and D is the desired delay. Doing so produces a peak in the fine structure of all impulse responses at D. Possible physiological sources of phase shift are discussed in Sec. III. The second assumption is that, after phase shifts, the channels can be summed. The entire operation (from input to sum) can be seen as a filter with impulse response

$$H_D(t) = \sum_k h_{k,D}(t), \tag{5}$$

where $h_{k,D}(t)$ is the phase-shifted impulse response of channel k. Figure 3(a) illustrates $H_D(t)$ for values of D ranging from 1 to 40 ms in 1 ms steps. Each response consists of a

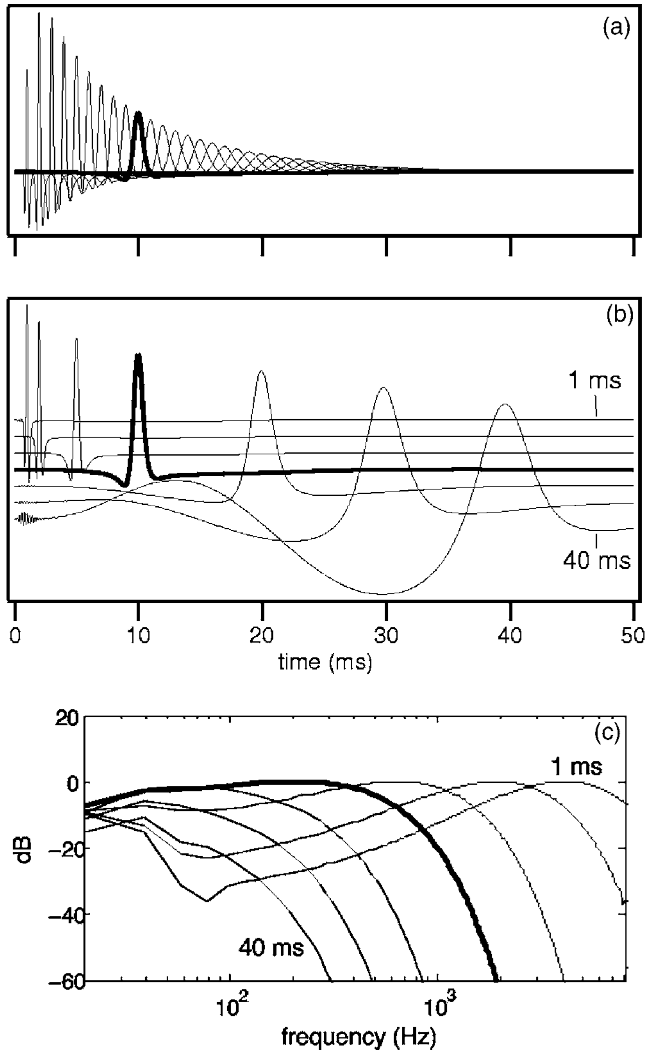

FIG. 3. (a) Impulse responses of gammatone-based synthetic delays of 1 to 40 ms in 1 ms steps. Bold: 10 ms. (b) Impulse responses for delays of 1, 2, 5, 10, 20, 30, and 40 ms, normalized for equal peak amplitude and offset vertically for clarity. Note the reduced amplitude and wider shape of impulse responses at longer delays. (c) Magnitude transfer functions of the same delays. Note the reduced amplitude and sharp low-pass characteristic of larger delays, and the more gradual high-pass characteristic of shorter delays.

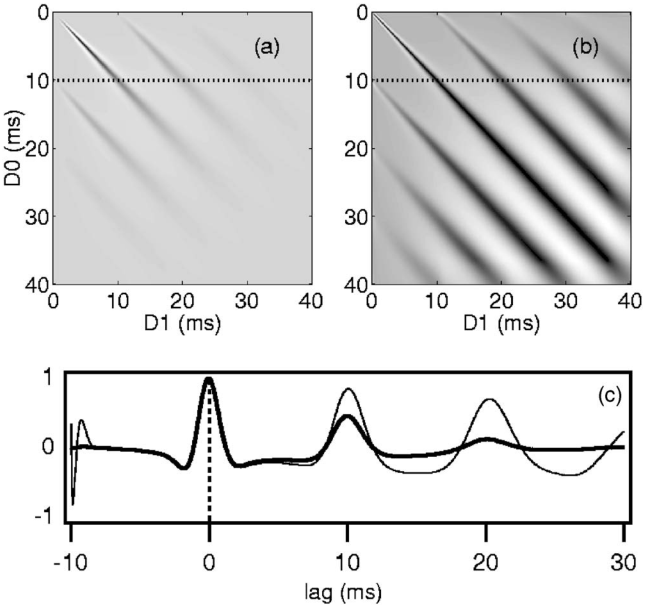

FIG. 4. (a) Cross product between wave forms delayed by $D0$ and $D1$, in response to a 100 Hz pulse train. (b) Same, normalized (see text), (c) Thick line: cross product plotted as a function of $D=D1-D0$, for $D0=10$ ms [dotted line in (a)]. This approximates the ACF. Thin line: same, normalized [dotted line in (b)].

pulse that appears to be delayed by D. Such delays will be termed *synthetic* to distinguish them from physical delays (e.g., neural transmission delays). Responses are indeed delayed, but two observations must be made. First, pulse amplitudes are smaller for larger D. Second, pulse shapes are not all the same, being wider for larger D. This is obvious in Fig. 3(b) where selected responses have been normalized for equal peak amplitude, and offset vertically for clarity. Figure 3(c) shows magnitude transfer functions of these selected synthetic delays. Longer delays have a relatively severe low-pass characteristic, and shorter delays a more gradual high-pass characteristic. This can be understood by noting that synthetic delays are restricted to the extent of impulse responses of the filter bank [Fig. 1(e)]. Responses within high-frequency channels fade quickly [Fig. 2(b)] so long delays are devoid of high frequencies. Conversely, the late onset of low frequency channels causes short delays to lack low frequencies. Details about simulations are given in the Appendix .

B. Using synthetic delays to produce autocorrelation

Synthetic delays may be used to approximate an autocorrelation function (ACF). The running ACF of a wave form may be defined as the cross product between the delayed wave form $x(t+D)$ and the undelayed $x(t)$

$$r_t(D) = (1/W) \sum_{j=t+1}^{t+W} x(j)x(j+D), \tag{6}$$

where W is the size of the integration window (supposed square for simplicity) and the subscript t indicates the time at which the calculation is made. We wish to implement Eq. (6) using the synthetic delay mechanism to obtain the delayed and undelayed terms. A complication is that we cannot set $D=0$ because all gammatone impulse responses are initially zero, and we do not have access to the original acoustic wave form. What we can do instead is calculate the cross product between two wave forms synthetically delayed by D_0 and D_1 such that $D_1-D_0=D$

$$r_t(D_0,D_1) = (1/W) \sum_{j=t+1}^{t+W} x_{D_0}(j)x_{D_1}(j), \tag{7}$$

where $x_{D_0}(t)$ and $x_{D_1}(t)$ are obtained by convolving $x(t)$ with $H_{D_0}(t)$ and $H_{D_1}(t)$, respectively. Figure 4(a) displays this cross product as a function of D_0 and D_1 for a stimulus consisting of a 100 Hz pulse train. Large values map to dark. The pattern consists of diagonal “stripes” with a spacing that reflects the stimulus period, 10 ms. An approximation of the ACF can be obtained by taking a section at any nonzero value of D_0 [Fig. 4(c), thick line].

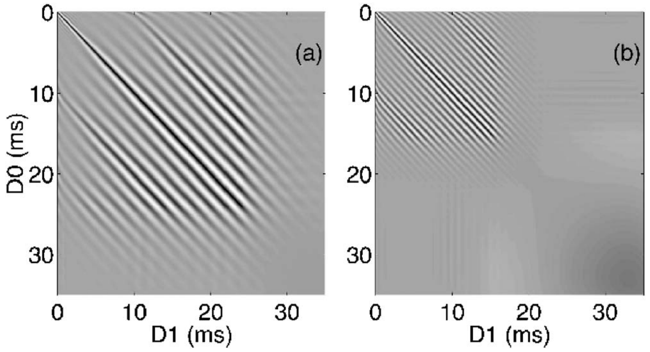

FIG. 5. Same as Figs. 4(a) and 4(b) for 100 Hz pulse trains high-pass filtered from 500 Hz (a) or 1000 Hz (b). The upper left-hand corner of each graph reflects stimulus periodicities (mainly the fundamental and component nearest cutoff).

The pattern fades for large values of D_0 and D_1 as a result of two factors: greater attenuation at longer delays [Fig. 3(a)] and delay-dependent spectral distortion [Fig. 3(c)] that reduces the similarity between $x_{D_0}(t)$ and $x_{D_1}(t)$. Effects of the first factor may be countered by normalizing the cross product

$$r'_t(D_0,D_1) = r_t(D_0,D_1) \times \left[\sum_{j=t+1}^{t+W} x_{D_0}(j)^2\right]^{-1/2}\left[\sum_{j=t+1}^{t+W} x_{D_1}(j)^2\right]^{-1/2}. \tag{8}$$

The normalized cross product is shown in Fig. 4(b) for the same wave form as in Fig. 4(a). Normalization requires a large dynamic range to be effective; if it were limited (as might be the case for a physiological implementation), the pattern would be intermediate between Figs. 4(a) and 4(b).

The thin line of Fig. 4(c) shows the approximation of the ACF obtained after amplitude normalization: The amplitude decrease is less fast than without normalization (thick line), but higher-order peaks are still somewhat broader than low-order peaks.

Figure 5 shows the normalized cross-correlation pattern in response to a 100 Hz complex tone high-pass filtered at 500 Hz (a) or 1000 Hz (b). Two things may be noted. First, periodicity-related structure is restricted to the upper left corner and does not extend beyond about 30 ms (a) or 15 ms (b). Second, the 10 ms fundamental pattern is overlayed by components near the cutoff: 2 ms in (a) and 1 ms in (b). These effects can be understood from the low-pass characteristic of most delays [Fig. 2(c)]. This illustrates once again an important property of synthetic delays: the range of delays that they offer is *frequency dependent*.

To summarize, relatively large "synthetic" delays may be produced by cross-channel phase interaction at the output of a cochlear model. Their size is limited by the CF-dependent duration of impulse responses: the longest delays are available only at low frequencies. The auditory system might use such delays to implement various processing mechanisms including autocorrelation. The rest of this paper discusses this idea in greater detail.

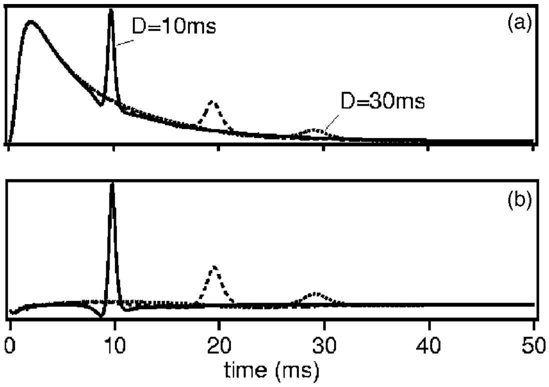

FIG. 6. Effects of non linear transduction. (a) Impulse responses for synthetic delays of 10, 20, and 30 ms in the case that filter outputs are processed by half-wave rectification. (b) Same, but half-wave rectification is followed by a high-pass "unsmoothing" filter.

III. ISSUES

This section examines several issues raised by the synthetic delay hypothesis. The purpose is to identify important issues but not necessarily resolve them.

A. Nonlinearities

Linear processing was assumed up to this point, but several stages of auditory processing are nonlinear. Hair cell transduction is corresponds roughly to half-wave rectification. Figure 6(a) illustrates the effect of introducing half-wave rectification before summation. Impulse responses are now dominated by a prominent initial portion that is common to all delays. However, Fig. 6(b) shows the result of high-pass filtering the half-wave rectified wave form in each channel (by subtracting the same wave form smoothed by convolution with a triangular window with a span of twice the characteristic period). Differences with Fig. 3 are now minimal. Such high-pass filtering might be obtained by a combination of fast excitatory and smoothed inhibitory inputs to a neuron, or by particular membrane properties (Kalluri and Delgutte, 2001).

Another source of nonlinearity is cochlear mechanics. Frequency and phase characteristics are known to change with level (Robles and Ruggero, 2001). If this were to result in level-dependent synthetic delays, the delays would be less useful. However measurements suggest that level affects mainly the relative *amplitudes* of the early and later segments of a chirp-shaped impulse response, leaving the temporal structure (position of peaks and zero crossings) invariant (Carney *et al.*, 1999; Robles and Ruggero, 2001; Shera, 2001). This in turn insures level invariance of synthetic delays. Shera (2001) commented that this property puts strong constraints on the mechanics and biophysical properties of the cochlea. We may note that it is a prerequisite to produce level-independent synthetic delays.

B. Loss of synchrony

The synthetic delay mechanism depends on precise temporal coding, which is known to degrade rapidly beyond

2.5–5 kHz in cat (Johnson, 1980). In man the limit is unknown. Stimulus components above that limit cannot be delayed.

C. Phase shifts

In our simulations, we implemented phase shifts by adjusting the phase term of the gammatone impulse response [Eq. (3)]. In the auditory system, a phase shift within $[0,2\pi]$ could be approximated by a physical delay within $[0,1/f]$.[2] This might seem to defeat the purpose of the model; the point is that these physical delays are shorter than the synthetic delays that they produce when combined across channels. The required range is further reduced by a factor of 2 if subtraction is allowed (e.g., inhibitory interaction). Phase shift may also be implemented as a weighted sum of two signals in quadrature phase, as suggested for binaural interaction by McAlpine *et al.* (2001). The appeal of this hypothesis is that the same set of signals would serve multiple purposes, and that weights may be easier to tune than physical delays. Also, only one physical delay ($1/4f$) is needed to produce the quadrature signal, rather than a range of delays. A third conceivable source of phase shifts is basilar membrane propagation (Shamma, 1985a; van der Heijden and Joris, 2005). Our model requires *independent* control of phase and CF whereas these factors presumably covary along the basilar membrane (BM), but the redundancy of overlapping channels might nevertheless allow BM phase shifts to be exploited.

D. Cross-channel summation

The neural equivalent of summation [Eq. (5)] would involve cross-frequency convergence of phase-locked inputs. This may occur at several stages within the auditory system, as reviewed further in Sec. V. The delayed pattern is supposed to be temporally accurate, and this narrows the options down to stages that have a synchronized *output* in addition to input. A possible way to relax this requirement is discussed below.

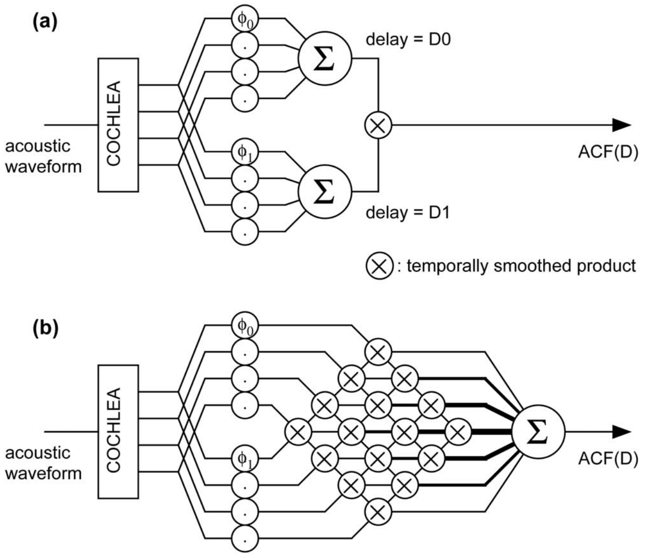

FIG. 7. (a) Operations required to produce one delay of the autocorrelation function. (b) Functionally equivalent circuit. The first circuit requires two stages of accurate temporal processing (the product is applied to fast-varying sums). The second needs only one stage (the sum is applied to temporally smoothed products).

E. Two possible implementations of the ACF

A first implementation is schematized in Fig. 7(a). It approximates the ACF for lag $D=D_1-D_0$. Each branch requires a cascade of two neurons, the first with additive properties and the second with coincidence-counting properties. These two steps might instead be combined within the dendritic field of a single neuron with appropriate properties. Agmon-Snir *et al.* (1998) give an example of complex dendrite-based computations. Calculating the ACF over a range of lags requires either an array of such phase-shift-and-sum circuits, or else a single circuit but with a tuning mechanism.

A second possible implementation is schematized in Fig. 7(b). To understand how it works, note that the right hand side of Eq. (7) can be expanded and rearranged

$$r_t(D_0,D_1) = \sum_{k,k'} r_{k,k',t}(D_0,D_1), \tag{9}$$

where

$$r_{k,k',t}(D_0,D_1) = (1/W)\sum_{j=t+1}^{t+W} x_{k,D_0}(j)x_{k',D_1}(j) \tag{10}$$

is the cross product between channel k phase shifted according to the formula for D_0, and channel k' phase shifted according to the formula for D_1. Equation (9) replaces the temporally smoothed product of fast-varying sums of Eq. (7) by a slow-varying sum of temporally smoothed products. The result is the same.

Several things make this second formulation attractive. First, a physiological implementation needs to maintain synchrony only up to the input of the initial (cross-product) stage, and not over two synapses as in Fig. 7(a). Second, cross coincidence between channels of different CF has been proposed repeatedly to account for a range of tasks including pitch and loudness perception (Schroeder, 1977; Carney, 1990; Shamma, 1985b; Shamma *et al.*, 1988; Deng and Geisler, 1987; Loeb *et al.*, 1983; Shamma and Klein, 2000; Heinz, 2001b; Carney *et al.*, 2002). This second formulation fits with the notion that the auditory system uses multipurpose processing, rather than specialized modules for each task such as pitch perception, and it adds to the range of possible incarnations of a useful functional model such as autocorrelation.

F. Cancellation

Delays are also required by hypothetical cancellation mechanisms involved in binaural or periodicity-based processing and sound segregation (e.g., Durlach, 1963, de Cheveigné, 1993). It has been argued (de Cheveigné, 2001) that cancellation and correlation-based statistics can be derived one from the other. On the basis of that argument we may assume that the same mechanism that produces the ACF can also produce a cancellation-based statistic, and that both statistics could be based in part on inhibitory as well as excitatory interaction. In other words, synthetic delays can fulfill the needs of cancellation models.

G. Parameters

Each synthetic delay involves K phase parameters ϕ_k, where K is the number of cochlear channels. All these parameters are functions of a single parameter via Eq. (1), but one may wonder how the auditory system applies that equation, or at least discovers which parameter sets correspond to delays. One possibility is that learning is involved in the tradition of learning-based models (Licklider, 1959; Terhardt, 1974; Shamma and Klein, 2000; see de Cheveigné, 2005 for a review). This issue is too complex to be further addressed within the scope of this paper.

To summarize, the synthetic delay model raises several issues that need to addressed before the model is deemed plausible. The next section examines how a model based on synthetic delays might account for pitch perception.

IV. SYNTHETIC DELAYS FOR PITCH PERCEPTION

In his model of pitch, Licklider (1951) proposed that spike trains from the cochlea were processed by a neural network consisting of a series of synaptic delays and coincidence neurons. Later models, such as Meddis and Hewitt (1991a,b), are less specific but assume nevertheless that delays are produced within the auditory system (by synaptic transmission, axonal conduction, or rebound from inhibition) with durations sufficient to calculate an ACF over the range necessary to for pitch perception. ACFs calculated within each CF channel are then summed to produce a *summary autocorrelation function* (SACF) from which the period is derived (Meddis and Hewitt, 1991a, b). Two major objections have been made to the AC model. The first is the lack of evidence for neural delays, at least at stages where phase locking is present. The second is that it works *too* well: It can handle any periodic wave form, and thus predicts that any periodic sound should evoke pitch. This argues against it, as pitch has limited regions of existence or salience. The synthetic delay model may help to address the second objection as well as the first.

Synthetic delays have two major limitations: (1) they cannot be applied to components for which there is no phase locking, and (2) they cannot exceed the impulse response duration of cochlear filters activated by the signal to be delayed. The first limit fits with the observation that periodicity pitch exists only for stimuli with components that allow phase locking (Moore, 2003, p. 215).[3] The second limit implies that period measurements are restricted to periods shorter than the impulse response of filters that respond to components of the tone. Such is the case if components are more widely spaced than filter bandwidths, that is, if they are *resolved*. Synthetic delays thus give the autocorrelation model a property that is usually associated with pattern-matching models. Data that show that resolvability is important for pitch (or other tasks) can potentially be explained by the use of synthetic delays.

Figure 8(a), dotted line, shows the latency of the gammatone envelope peak as a function of CF, in units of characteristic period (CP, inverse of CF). Latencies of the skirts at −20 dB are plotted as thick lines, and at −40 and −60 dB as thinner lines. If the dynamic range of processing were 20 dB, the useful duration would be the interval between thick lines, replotted as a thick line in Fig. 8(b). Superimposed upon that plot are data from several studies that measured pitch discrimination. Symbols indicate the rank and frequency of the lowest component of stimuli used in those studies. Large triangles are the resolved (full) and unresolved (open) conditions of Carlyon and Shackleton (1994). Small symbols are for other studies that measured pitch discrimination thresholds (Houtsma and Smurzynski, 1990; Krumbholz, Patterson, and Pressnitzer, 2000; Kaernbach and Bering, 2001; Bernstein and Oxenham, 2003). For each study, full symbols represent conditions for which pitch thresholds were low (below the geometric mean over conditions) and open symbols conditions for which they were high.

The 20 dB line separates the two populations quite neatly: good performance below, and poor above. This line is a contour of constant resolvability, and the data thus illustrate the well-known observation that pitch discrimination is better for stimuli with at least some resolved components. The line also marks the maximum delay that can be synthesized by the model, and the data thus equally support the hypothesis that synthetic delays are required for accurate pitch. This conclusion is based on the 20 dB value chosen for the dynamic range, and also on the gammatone model chosen to parametrize auditory filters.

Note that maximum delay *covaries* with resolvability but does not depend upon it. Bernstein and Oxenham (2003) found that discrimination thresholds were not improved when odd partials were sent to one ear and even partials to the other. That manipulation increases interpartial spacing at each ear, and should improve discrimination if resolvability were the determining factor. Their data fit the hypothesis, put forward by Moore (2003) and recently explored by Bernstein and Oxenham (2005), that internal delays are limited to some *ad hoc* value that depends on the spectral range occupied by the stimulus. For the synthetic delay model this limit is an emergent property. The model is consistent with other aspects of pitch that are hard to explain on the basis of autocorrelation. Bernstein and Oxenham (2005) found that the transition from good to poor pitch discrimination occurs at a lower harmonic rank at high stimulus levels, for which cochlear filters are wider and impulse responses presumably shorter. Supposing that pure tone discrimination involves higher order peaks of the ACF (de Cheveigné, 1989; Slaney, 1990), the same explanation might account for elevated pure tone thresholds at high levels (Bernstein and Oxenham, 2005). Synthetic delays require intact cochlear filters. Absence of filters in cochlear implantees might explain their lack of accurate pitch perception, despite the precise temporal patterns of nerve discharge produced by electrical stimulation.

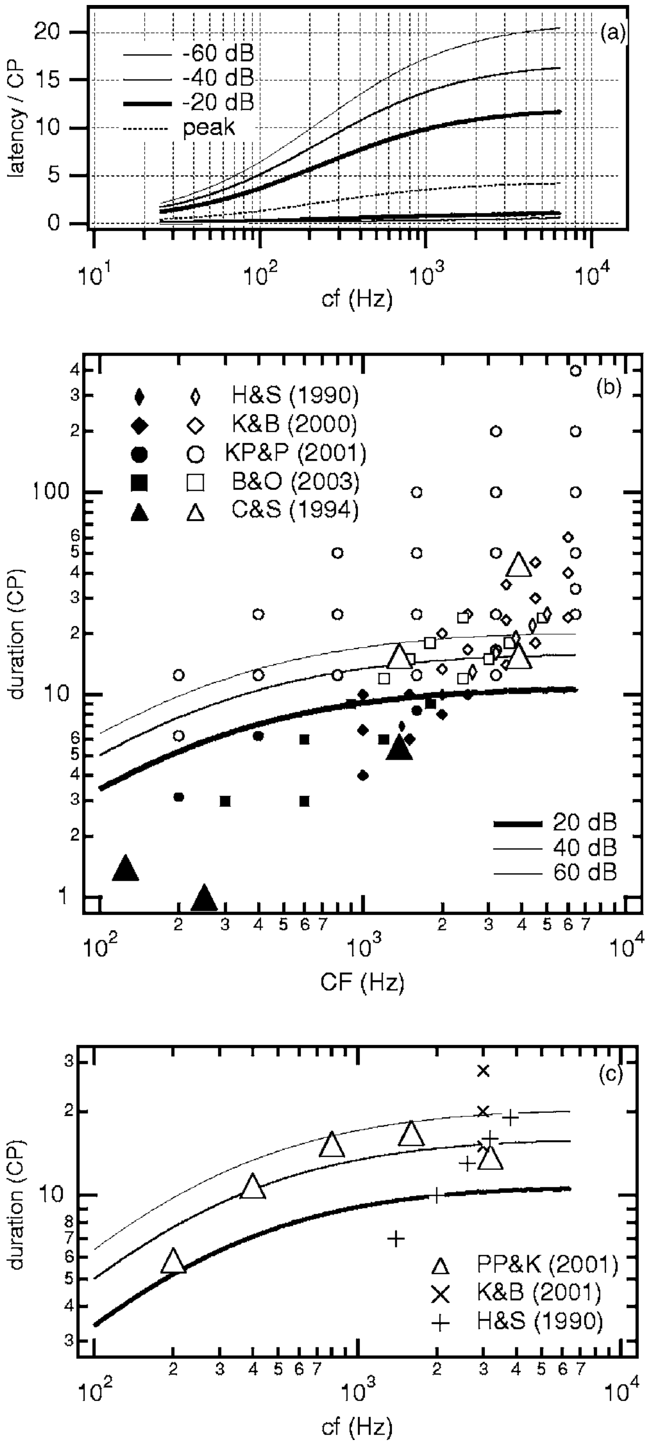

FIG. 8. (a) Latencies of the peak (dotted line) and onset and offset (full lines) of gammatone filter impulse responses in units of characteristic period (1/CF). The useful duration is the interval between onset and offset. (b) Thick line: useful duration of the gammatone impulse response given a dynamic range of 20 dB. Thin lines: same for 40 and 60 dB. Symbols represent the rank of the lowest component of stimuli used in pitch discrimination experiments (Houtsma and Smurzynski, 1990; Carlyon and Shackleton, 1994; Krumbholz, Patterson, and Pressnitzer, 2000; Kaernbach and Bering, 2001; Bernstein and Oxenham, 2003). For each study, open (respectively, full) symbols represent conditions for which threshold was above (respectively, below) the geometric mean over conditions. The transition from good to poor discrimination in most of these studies was typically abrupt and large (up to an order of magnitude). (c) Lines are as in (b). Crosses represent the rank of the lowest component of stimuli for which musical tasks were performed at better than chance level (Houtsma and Smurzynski, 1990; Kaernbach and Bering, 2001). The triangles are the lower limit of melodic pitch as determined by Pressnitzer *et al.* (2001b).

A difficulty, shared with pattern matching models, is that stimuli with unresolved components may still evoke a pitch, albeit with weak salience and poor accuracy. Indeed, the studies shown in Fig. 8(c) report threshold measurements for ranks of up to 400. The pitch of such stimuli is weak but may still be musical. Houtsma and Smurzynski (1990) and Kaernbach and Bering (2001) found better-than-chance interval recognition for stimuli with ranks of up to 19 and 28, respectively. Pressnitzer *et al.* (2001a) likewise found a lower limit of melodic pitch that was well outside the limits of resolvability [Fig. 8(c)].

It is worth noting, however, that the limit plotted in Fig. 8 is conditional on the gamma-tone model chosen to represent cochlear filters and on its parameters. From human otoacoustic emission and behavioral data on nonsimultaneous masking, Shera *et al.* (2002) and Oxenham and Shera (2003) argued that filters might be twice as narrow as found by Moore and Glasberg (1983), and thus their impulse responses twice as long. If that conclusion can be extended to the context of complex tones, the limit of Fig. 8 might shift to higher ranks. Measurements of basilar membrane motion or auditory nerve responses reveal complex, multiple-spindle-shaped click or revcor responses (Recio *et al.*, 1998; Recio and Rhode, 2000; Lin and Guinan, 2000, 2004) that do not conform to the gammatone model. In at least one case the response extended in time as far as the 90th cycle (Recio *et al.*, 1998, Fig. 4). Thus, it is possible that the synthetic mechanism could operate beyond the limit shown in Fig. 8.

However, this explanation cannot account for pitch perception with electrical stimulation (Pilj and Scharz, 1995). For that, one would need to postulate some other source of delay, for example, rebound from inhibition, or axonal conduction. This "dual-delay" hypothesis, which could also account for the pitch of unresolved stimuli, resembles the "dual pitch mechanism" hypothesis (Houtsma and Smurzynski, 1990; Carlyon and Shackleton, 1994). It differs in that the hypothetical pitch mechanism that follows the delays could be unitary, avoiding the need for a "translation" mechanism to convert outputs of different mechanisms to a common ground. Evidence for "translation noise" was found by Carlyon and Shackleton (1994), but their interpretation has recently been challenged (Micheyl and Oxenham, 2004; Gockel *et al.*, 2004; Oxenham *et al.*, 2005).

To summarize, the synthetic delay mechanism can be used to implement a delay-based pitch model such as autocorrelation. The availability of period-size synthetic delays coincides with component resolvability. Data that show the importance of resolvability for pitch are consistent with the hypothesis that synthetic delays are involved. However pitch is evoked also by stimuli that are not resolvable, and in subjects that lack peripheral filters. Another source of delay is needed to handle cases for which the synthetic delay mechanism is unavailable.

V. PHYSIOLOGICAL EVIDENCE

A way to evaluate the synthetic delay hypothesis is to find anatomical or physiological evidence for or against the ingredients that it assumes, or the activity that it would im-

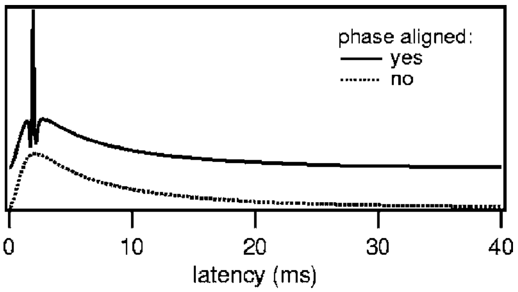

FIG. 9. Sum across auditory fibers of modeled firing probabilities in response to a click, with (full) and without (dashed) phase alignment. Plots are shifted vertically for clarity.

ply. These vary according to whether we assume the standard implementation of Fig. 7(a), or the alternative implementation of Fig. 7(b).

The implementation of Fig. 7(a) leads us to expect to observe phase-locked, delayed responses somewhere within the auditory system. Lack of evidence for them is as damning as for other hypothetical sources of delay such as axonal conduction, although it is possible that they exist but in a form that is hard to observe (for example, within thin axones or dendrites). In contrast, the implementation of Fig. 7(b) does not imply observable delays. In that implementation, the ACF is "assembled" from between-channel cross-correlation terms. Both implementations lead us to expect responses tuned to periodicity pitch. The lack of period-selective responses is a problem for any model of pitch, including pattern matching (but see Bendor and Wang, 2005).

Both implementations require cross-frequency convergence of phase-locked inputs. This is known to occur in the cochlear nucleus, for example, in octopus (onset-I) and D-stellate (chopper-C) cells (Jiang *et al.*, 1996; Oertel *et al.*, 2000). It could also, occur at higher stages that receive accurate temporal patterns relayed by primarylike cells, such as the superior olivary complex, nuclei of the lateral lemniscus, or inferior colliculus. Both implementations assume linear summation, of fast patterns in Fig. 7(a) or slow in Fig. 7(b). Accurate linear summation has been observed in cultured pyramidal cells (Cash and Yuste, 1998). The implementation of Fig. 7(b) assumes cross coincidence between channels with different CF, an assumption made by many models (e.g., Carney, 1990).

Both implementations assume that the *phase* of each cochlear channel can be manipulated. Onset-I and chopper-C cells of the cochlear nucleus receive input from AN fibers over a wide range of CFs (Jiang *et al.*, 1996; Oertel *et al.*, 2000). They respond to clicks with little temporal jitter, despite the fact that activity within fibers that feed them is presumably staggered in time due to phase dispersion along the basilar membrane. Response to a complex in random phase is less good than sine or cosine phase, implying that cross-frequency phase alignment is critical (Palmer and Winter, 1996; Evans and Zhao, 1997; Shofner, 1999; Winter and Palmer, 1995). Figure 9 (dashed line) shows an estimate of nerve firing probability pooled across the entire AN, as might drive one of these onset cells (based on a gammatone model of peripheral filtering). With no phase alignment the pattern is relatively wide, but aligning the phases makes it temporally much sharper (full line). Octopus cells require multiple excitatory postsynaptic potentials (EPSPs) to occur within a very short window (about 1 ms) to fire (Oertel *et al.*, 2000), a condition obviously easier to meet if inputs are phase aligned. In the onset cell model of Kalluri and Delgutte (2001), realistic entrainment required that inputs span less than 1.5 kHz, an unrealistically small range (Jiang *et al.*, 1996). That constraint could have been relaxed if phases were aligned. Observation of a *range* of phases for a given CF would favor the model. Indeed, McAlpine *et al.* (2001) report responses in quadrature phase from which other phase relations may be synthesized.

Winter and Palmer (1995) reported first spike latencies for onset cells between 0.5 and 5 ms, but these were level dependent. Behrend *et al.* (2002) found cells in the gerbil superior paraolivary nucleus (SPN) with temporally accurate onset responses (jitter smaller than 100 μs for a subset of cells) and a range of latencies (0.9 to 11.2 ms). Some were phase locked to amplitude modulation up to 1 kHz, and most had CFs below 6 kHz and relatively broad tuning. SPN receives inhibitory input from medial nucleus of trapezoid body (MNTB) and excitatory input from multipolar and octopus cells in CN, that both integrate inputs over a range of CFs, and the dendritic fields of its neurons span a wide range (Dehmel *et al.*, 2002). However such evidence for systematic delays, as might be produced by the implementation of Fig. 7(a) (or other forms of neural delay) remains fragmentary. To summarize, ingredients required by the synthetic delay model are available even if explicit evidence for the model is lacking.

VI. DELAY, PHASE AND COCHLEAR FILTERING

The concept of the ear as a Fourier transformer dates to von Helmholtz (1877), who noted that we are deaf to the relative phases of low-order partials. It was subsequently taken for granted that phase information is discarded, but Huggins and Licklider (1951; Huggins, 1952) pointed out that frequency analysis can be enhanced by phase interaction between channels of a filterbank. Similar ideas are embodied in the lateral inhibitory network (LIN) of Shamma (1985b) which involves subtraction between adjacent channels to sharpen selectivity, or the model of Deng and Geisler (1987) involving multiplicative interaction. Carney (1990) searched for evidence of such interaction in the cochlear nucleus, and Heinz *et al.* (2001a, b) and Carney *et al.* (2002) suggested that it can explain loudness perception and tone detection in noise. In another twist on pitch perception, Shamma and Klein (2000) used cross coincidence to produce harmonic templates as required by pattern-matching models. Together, these phase-sensitive mechanisms form a richer set than those based only on magnitude spectral patterns. In the models of Jeffress (1948) and Licklider (1951), coincidence detectors are preceded by neural delays, that are also required by equalization-cancellation (Durlach, 1963), harmonic cancellation (de Cheveigné, 1993) or strobed temporal integration (Patterson *et al.*, 1992). Delays were initially assumed to be neural, but basilar membrane propagation delay has also been proposed for binaural (Schroeder, 1977; Shamma *et al.*, 1988, 1989; Shamma, 2001; Joris *et al.*, 2005) and monaural processing (Hurst, 1895; Loeb, 1983). For the narrow-band signal that exists within a cochlear channel, delay is equivalent to *phase shift* and can be implemented as a weighted sum of two terms in quadrature phase. McAlpine *et al.* (2001) argued that such a mechanism might underly localization in the guinea pig, in place of Jeffress's delay-based mechanism. The synthetic delay model draws upon all these earlier ideas and models.

Our model involves peripheral filtering and phase manipulation. It could conceivably be reformulated in terms of pattern matching on a complex spectral representation (involving fine structure) as in Shamma and Klein (2000). Our formulation in terms of delay would nevertheless remain useful, at least for pedagogical reasons. It is not equivalent to a pattern-matching model based on a magnitude spectrum (e.g., a rate-based tonotopic pattern).

VII. CONCLUSION

1. Cross-channel phase interaction between cochlear channels can produce "synthetic" delays of up to about 30 ms, as a result of interference between phase-shifted impulse responses of cochlear filters. Synthetic delays offer an alternative to neural delays in models of auditory processing.
2. Synthetic delays are limited by the finite duration of cochlear filter impulse responses, that tend to vary inversely with bandwidth: filters are wider at high CFs, so maximum delay is shorter in high-CF than in low-CF channels. The *ratio* of maximum delay to characteristic period (1/CF) is however greater in high-CF than in low-CF channels.
3. Synthetic delays are contingent on phase locking, and ineffective for stimuli with components beyond about 2.5–5 kHz (if this happens to be the limit for humans).
4. Synthetic delays may be used in models of pitch based on autocorrelation, or models of segregation based on cancellation. Limits on delay duration and synchrony impose performance limits for these models. Limits for pitch are roughly consistent with those observed behaviorally, in particular (a) a lower limit of pitch of about 30 Hz for wide-band stimuli, (b) the increase of this limit when stimuli are high-pass filtered, (c) poor pitch discrimination for stimuli containing only harmonics with ranks >10, and more generally (d) the apparent dependency of accurate pitch on *resolvability* of stimulus components. The model is also consistent with the absence of accurate pitch perception for electrical stimulation.
5. The model can be formulated so as to use an initial stage of processing based on between-channel cross correlation that might be shared with other recent models of auditory processing that postulate such interaction.
6. In order to explain the weak pitch of stimuli with "unresolved" harmonics, it may be necessary to postulate an additional, presumably less accurate, source of delay such as produced by axonal conduction or rebound from inhibition.

ACKNOWLEDGMENTS

The authors wish to thank Maria Chait, Christian Kaernbach, Shihab Shamma, Mounya Elhilali, Andrew Oxenham, and Armin Kohlrausch for insightful comments on this manuscript. Katharine Earp provided last minute inspiration.

APPENDIX. SIMULATION DETAILS

Simulations of synthetic delays were performed with a filterbank comprising 2000 channels uniformly spaced on an ERB-rate scale from 25 to 10 000 Hz. Gammatone filters were implemented in the time domain using the "Auditory Toolbox" of Slaney (1993), or in the frequency domain based on the formula

$$Gf = (n-1)\,!\,(1/2)([\cos(\phi) - i\sin(\phi)]\{b/2\pi[i(f+f_c) + b]\}^n + [\cos(\phi) + i\sin(\phi)]\{b/(2\pi[i(f-f_c)+b])\}^n) \tag{A1}$$

Roll off of synchrony at high frequencies was simulated by applying to the stimulus wave form a low-pass filter consisting of seven cascaded first-order low-pass filters with cutoffs at 4800 Hz. This filter produces a −3 dB cutoff at 2500 Hz followed by a 100 dB/decade rolloff that simulates loss of AN synchrony (Heinz *et al.*, 2001a).

---

[1] Other shortcomings of autocorrelation models have been put forward (Kaernbach and Demany, 1998; Pressnitzer *et al.*, 2001b, 2004) but they do not relate to the issue of delays, so they will not be discussed in this paper.

[2] Delay shifts both the envelope and the fine structure, whereas phase shift affects only the latter. Simulations (not shown) show that the synthetic delay mechanism is nevertheless effective when phase shifts are produced in this way.

[3] Missing-fundamental stimuli have a low pitch up to about 1400 Hz, at which point the third harmonic, 4200 Hz, falls close to the phase locking limit.

Agmon-Snir, H., Carr, C., and Rinzel, J. (**1998**). "The role of dendrites in coincidence detection," Nature (London) **393**, 268–272.

Behrend, O., Brand, A., Kapfer, C., and Grothe, B. (**2002**). "Auditory response properties in the superior paraolivary nucleus of the gerbil," J. Neurophysiol. **87**, 2915–2928.

Bendor, D., and Wang, X. (**2005**). "The neuronal representation of pitch in the auditory cortex," Nature (London) **436**, 1161–1165.

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

Bernstein, J. G. W., and Oxenham, A. (**2005**). "An autocorrelation model with place dependence to account for the effect of harmonic number on fundamental frequency discrimination," J. Acoust. Soc. Am. **117**, 3816–3831.

Blodgett, H. C., Wilbanks, W. A., and Jeffress, L. A. (**1956**). "Effect of large interaural time differences upon the judgment of sidedness," J. Acoust. Soc. Am. **28**, 639–643.

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

Cariani, P. A., 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. P. (**1998**). "Comments on "A unitary model of pitch perception"

[J. Acoust. Soc. Am. **102**, 1811–1820 (1997)]," 1998). "Comments on "A unitary model of pitch perception" J. Acoust. Soc. Am. **104**, 1118–1121.

Carlyon, R. P., and Shackleton, T. M. (**1994**). "Comparing the fundamental frequencies of resolved and unresolved harmonics: evidence for two pitch mechanisms?," J. Acoust. Soc. Am. **95**, 3541–3554.

Carney, L. (**1990**). "Sensitivities of cells in anteroventral cochlear nucleus of cat to spatiotemporal discharge patterns across primary afferents," J. Neurophysiol. **64**, 437–456.

Carney, L. H., Heinz, M. G., Evilsizer, M. E., Gilkey, R. H., and Colburn, H. S. (**2002**). "Auditory phase opponency: A temporal model for masked detection at low frequencies," Acta. Acust. Acust. **88**, 334–347.

Carney, L., McDuffy, M. J., and Shekhter, I. (**1999**). "Frequency glides in the impulse responses of auditory-nerve fibers," J. Acoust. Soc. Am. **105**, 2384–2391.

Carney, H., and Yin, T. C. T. (**1988**). "Temporal coding of resonances by low-frequency auditory nerve fibers: Single fiber responses and a population model," J. Neurophysiol. **60**, 1653–1677.

Cash, S., and Yuste, R. (**1998**). "Input summation by cultured pyramidal neuron is linear and position independent," J. Neurosci. **18**, 10–15.

de Boer, E. (**1956**). "On the "residue" in hearing," University of Amsterdam, unpublished doctoral dissertation.

de Cheveigné, A. (**1989**). "Pitch and the narrowed autocoincidence histogram," *Proc. ICMPC*, Kyoto, pp. 67–70 (The Japanese Society of Music Perception and Cognition).

de Cheveigné, A. (**1993**). "Separation of concurrent harmonic sounds: Fundamental frequency estimation and a time-domain cancellation model of auditory processing," J. Acoust. Soc. Am. **93**, 3271–3290.

de Cheveigné, A. (**1997**). "Concurrent vowel identification III: A neural model of harmonic interference cancellation," J. Acoust. Soc. Am. **101**, 2857–2865.

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

de Cheveigné, A. (**1999**). "Pitch shifts of mistuned partials: A time-domain model," J. Acoust. Soc. Am. **106**, 887–897.

de Cheveigné, A. (**2000**). "A model of the perceptual asymmetry between peaks and troughs of frequency modulation," J. Acoust. Soc. Am. **107**, 2645–2656.

de Cheveigné, A. (**2001**). "Correlation Network model of auditory processing," *Proc. Workshop on Consistent & Reliable Acoustic Cues for Sound Analysis* (Aalborg, Danmark).

de Cheveigné, A. (**2005**). "Pitch perception models," in *Pitch*, edited by C. Plack and A. Oxenham Springe-Verlag, New York, pp. 169–233.

Dehmel, S., Kopp-Scheinpflug, C., Dörrscheidt, G. J., and Rübsamen, R. (**2002**). "Electrophysiological characterization of the superior paraolivary nucleus in the mongolian gerbil," Hear. Res. **172**, 18–36.

Deng, L., and Geisler, C. D. (**1987**). "A composite auditory model for processing speech sounds," J. Acoust. Soc. Am. **81**, 2001–2012.

Durlach, N. I. (**1963**). "Equalization and cancellation theory of binaural masking-level differences," J. Acoust. Soc. Am. **35**, 1206–1218.

Evans, E. F., and Zhao, W. (**1997**). "Periodicity coding of the fundamental frequency of harmonic complexes: Physiological and pharmacological study of onset units in the ventral cochlear nucleus," in *Psychophysical and Physiological Advances in Hearing*, edited by A. Palmer, A. Rees, A. Q. Summerfield, and R. Meddis (London, Whurr), pp. 186–194.

Gockel, H., Carlyon, R. P., and Plack, C. J. (**2004**). "Across-frequency interference in fundamental frequency discrimination: Questioning evidence for two pitch mechanisms," J. Acoust. Soc. Am. **116**, 1092–1104.

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

Heinz, M. G., Colburn, H. S., and Carney, L. H. (**2001a**). "Evaluating auditory performance limits: I. One-parameter discrimination using a computational model for the auditory nerve," Neural Comput. **13**, 2273–2316.

Heinz, M. G., Colburn, H. S., and Carney, L. H. (**2001b**). "Rate and timing cues associated with the cochlear amplifier: Level discrimination based on monaural cross-frequency coincidence detection," J. Acoust. Soc. Am. **110**, 2065–2084.

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

Huggins, W. H. (**1952**). "A phase principle for complex-frequency analysis and its implications in auditory theory," J. Acoust. Soc. Am. **4**, 582–589.

Huggins, W. H., and Licklider, J. C. R. (**1951**). "Place mechanisms of auditory frequency analysis," J. Acoust. Soc. Am. **23**, 290–299.

Hurst, C. H. (**1895**). "A new theory of hearing," Proc. Trans. Liverpool Biol. Soc. **9**, 321–353 (and plate XX).

Jeffress, L. A. (**1948**). "A place theory of sound localization," J. Comp. Physiol. Psychol. **41**, 35–39.

Jiang, D., Palmer, A., and Winter, I. M. (**1996**). "Frequency extent of two-tone facilitation in onset units in the ventral cochlear nucleus," J. Neurophysiol. **75**, 380–395.

Johnson, D. H. (**1980**). "The relationship between spike rate and synchrony in responses of auditory-nerve fibers to single tones," J. Acoust. Soc. Am. **68**, 1115–1122.

Joris, P. X., van der Heijden, M., Louage, D., Van de Sande, B., and Van Kerckhoven, C. (**2005**). "Dependence of binaural and cochlear "best delay" on characteristic frequency," in *Auditory Signal Processing: Physiology, Psychoacoustics, and Models*," edited by D. Pressnitzer, A. de Cheveigné, S. McAdams, and L. Collet (Springer, New York), pp. 478–484.

Kaernbach, C., and Bering, C. (**2001**). "Exploring the temporal mechanism involved in the pitch of unresolved harmonics," J. Acoust. Soc. Am. **110**, 1039–1048.

Kaernbach, C., and Demany, L. (**1998**). "Psychophysical evidence against the autocorrelation theory of pitch perception." J. Acoust. Soc. Am. **104**, 2298–2306.

Kalluri, S., and Delgutte, B. (**2001**). "Characteristics of cochlear nucleus onset units studied with a model," "*Computational Models of Auditory Function*," edited by S. Greenberg and M. Slaney (IOS Press, Amsterdam), pp. 29–48.

Krumbholz, K., Patterson, R. D., and Pressnitzer, D. (**2000**). "The lower limit of pitch as determined by rate discrimination," J. Acoust. Soc. Am. **108**, 1170–1180.

Licklider, J. C. R. (**1951**). "A duplex theory of pitch perception," Experientia **7**, 128–134.

Licklider, J. C. R. (**1959**). "Three auditory theories," "*Psychology, a Study of a Science*," edited by S. Koch (McGraw-Hill, New York), **I**, pp. 41–144.

Lin, T., and Guinan, J. J. (**2000**). "Auditory-nerve-fiber responses to high-level clicks: interference patterns indicate that excitation is due to the combination of multiple drives," J. Acoust. Soc. Am. **107**, 2615–2630.

Lin, T., and Guinan, J. J., Jr (**2004**). "Time-frequency analysis of auditory-nerve-fiber and basilar-membrane click responses reveal glide irregularities and noncharacteristic-frequency skirts," J. Acoust. Soc. Am. **116**, 405–416.

Loeb, G. E., White, M. W., and Merzenich, M. M. (**1983**). "Spatial cross-correlation - A proposed mechanism for acoustic pitch perception," Biol. Cybern. **47**, 149–163.

McAlpine, D., Jiang, D., and Palmer, A. (**2001**). "A neural code for low-frequency sound localization in mammals," Nat. Neurosci. **4**, 396–401.

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

Meddis, R., and Hewitt, M. J. (**1991b**). "Virtual pitch and phase sensitivity of a computer model of the auditory periphery. II: Phase sensitivity," J. Acoust. Soc. Am. **89**, 2883–2894.

Meddis, R., and O'Mard, L. (**1997**). "A unitary model of pitch perception," J. Acoust. Soc. Am. **102**, 1811–1820.

Micheyl, C., and Oxenham, A. (**2004**). "Sequential F0 comparisons between resolved and unresolved harmonics: No evidence for translation noise between two pitch mechanisms," J. Acoust. Soc. Am. **116**, 3038–3050.

Moore, B. C. J. (**2003**). *An Introduction to the Psychology of Hearing* (Academic Press, London).

Moore, B. C. J., and Glasberg, B. R. (**1983**). "Suggested formulae for calculating auditory-filter bandwidths and excitation patterns," J. Acoust. Soc. Am. **74**, 750–753.

Mossop, J. E., and Culling, J. F. (**1999**). "Lateralization of large interaural delays," J. Acoust. Soc. Am. **104**, 1574–1579.

Oertel, D., Bal, R., Gardner, S. M., Smith, P. H., and Joris, P. X. (**2000**). "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., pp. 11773–11779.

Oxenham, A., Bernstein, J. G. W., and Penagos, H. (**2004**). "Correct tonotopic representation is necessary for complex pitch perception," Proc. Natl. Acad. Sci. U.S.A. **101**, 1421–1425.

Oxenham, A., Bernstein, J. G., and Micheyl, C. (**2005**). "Pitch perception of complex tones within and across ears and frequency regions," *Auditory Signal Processing: Physiology, Psychoacoustics, and Models*, edited by D. Pressnitzer, A. de Cheveigné, S. McAdams, and L. Collet (Springer, New York), pp. 126–135.

Oxenham, A., and Shera, C. A. (**2003**). “Estimates of human cochlear tuning at low levels using forward and simultaneous masking,” J. Assoc. Res. Otolaryngol. **4**, 541–554.

Palmer, A. R., and Winter, I. M. (**1996**). “The temporal window of two-tone facilitation in onset units of the ventral cochlear nucleus,” J. Nurs. Meas **1**, 12–30.

Patterson, R. D., Allerhand, M., and Guiguère, C. (**1995**). “Time-domain modeling of peripheral auditory processing: Amodular architecture and a software platform,” J. Acoust. Soc. Am. **98**, 1890–1894.

Patterson, R. D., Robinson, K., Holdsworth, J., McKeown, D., Zhang, C., and Allerhand, M. (**1992**). “Complex sounds and auditory images,” *Auditory Physiology and Perception*, edited by Y. Cazals, K. Horner, and L. Demany (Pergamon Press, Oxford), pp. 429–446.

Pilj, S., and Scharz, D. W. F. (**1995**). “Melody recognition and musical interval perception by deaf subjects stimulated with electric pulse trains through single implant electrodes,” J. Acoust. Soc. Am. **98**, 886–895.

Pressnitzer, D., de Cheveigné, A., and Winter, I. M. (**2004**). “Physiological correlates of the perceptual pitch shift of sounds with similar waveform autocorrelation,” ARLO **5**, 1–6.

Pressnitzer, D., Patterson, R. D., and Krumbholz, K. (**2001a**). “The lower limit of melodic pitch,” J. Acoust. Soc. Am. **109**, 2074–2084.

Pressnitzer, D., Winter, I. M., and de Cheveigné, A. (**2001b**). “Perceptual pitch shift for sounds with similar waveform autocorrelation,” ARLO **3**, 1–6.

Recio, A., and Rhode, W. S. (**2000**). “Basilar membrane responses to broadband stimuli,” J. Acoust. Soc. Am. **108**, 2281–2298.

Recio, A., Rich, N. C., Narayan, S. S., and Ruggero, M. A. (**1998**). “Basilar-membrane responses to clicks at the base of the chinchilla cochlea,” J. Acoust. Soc. Am. **103**, 1972–1989.

Robles, L., and Ruggero, M. A. (**2001**). “Mechanics of the mammalian cochlea,” Physiol. Rev. **81**, 1305–1352.

Saberi, K., Takahashi, Y., Egnor, R., Farahbod, H., and Mazer, J. (**2001**). “Detection of large interaural delays and its implication for models of binaural interaction,” J. Assoc. Res. Otolaryngol. **3**, 80–88.

Schroeder, M. R. (**1977**). “New viewpoints in binaural interaction,” in *Psychophysics and Physiology of Hearing*,” edited by E. F. Evans and J. P. Wilson (Academic Press, London), pp. 455–467.

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

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

Shamma, S. A. (**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. (**2001**). “On the role of space and time in auditory processing,” Trends in Cognitive Science **5**, 340–348.

Shamma, S., N., S., and Gopalaswamy, P. (**1988**). “Binaural processing without neural delays,” in *Basic Issues in Hearing*, edited by H. Duifhuis, J. W. Horst, and H. P. Wit (Academic Press, London), pp. 135–143.

Shamma, S., and Klein, D. (**2000**). “The case of the missing pitch templates: How harmonic templates emerge in the early auditory system,” J. Acoust. Soc. Am. **107**, 2631–2644.

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

Shera, C. A. (**2001**). “Intensity-invariance of fine time structure in basilar-membrane click responses: Implications for cochlear mechanics,” J. Acoust. Soc. Am. **110**, 332–348.

Shera, C. A., Guinan, J. J., and Oxenham, A. J. (**2002**). “Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements,” Proc. Natl. Acad. Sci. U.S.A. **99**, 3318–3323.

Shofner, W. P. (**1999**). “Responses of cochlear nucleus units in the chinchilla to iterated rippled noises: Analysis of neural autocorrelograms,” J. Neurophysiol. **81**, 2662–2674.

Slaney, M. (**1990**). “A perceptual pitch detector,” *Proc. International Conference on Acoustics and Speech and Signal Processing (IEEE)*, pp. 357–360.

Slaney, M. (**1993**). “An efficient implementation of the Patterson-Holdsworth auditory filter bank,” Apple Computer Technical Report, 35.

Stern, R. M., and Shear, G. D. (**1996**). “Lateralization and detection of low-frequency binaural stimuli: Effects of distribution of internal delay,” J. Acoust. Soc. Am. **100**, 2278–2288.

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

van der Heijden, M., and Joris, P. X. (**2005**). “Reconstructing the traveling wave from auditory nerve responses,” in *Auditory Signal Processing: Psychophysics, Physiology and Modeling*,” edited by D. Pressnitzer, A. de Cheveigné, S. McAdams, and L. Collect (Springer-Verlag, New York), pp. 7–13.

von Helmholtz, H. L. F. (**1877**). *On the Sensations of Tone* (English translation A. J. Ellis, 1885, 1954), (Dover, New York).

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

Winter, I. M., and Palmer, A. (**1995**). “Level dependence of cochlear nucleus onset unit response and facilitation by second tones or broadband noise,” J. Neurophysiol. **73**, 141–159.