

# Cancellation model of pitch perception<sup>a)</sup>

Alain de Cheveigné

Laboratoire de Linguistique Formelle, CNRS/Université Paris 7, 2 place Jussieu, case 7003, 75251, Paris, France and ATR Human Information Processing Research Laboratories, 2-2 Hikaridai, Seikacho, Soraku-gun, Kyoto 619-02, Japan

(Received 7 May 1997; accepted for publication 13 October 1997)

A model of pitch perception is presented involving an array of delay lines and inhibitory gating neurons. In response to a periodic sound, a *minimum* appears in the pattern of outputs of the inhibitory neurons at a lag equal to the period of the sound. The position of this minimum is the cue to pitch. The model is similar to the autocorrelation model of pitch, multiplication being replaced by an operation similar to subtraction, and maxima by minima. The two models account for a wide class of pitch phenomena in very much the same way. The principal goal of this paper is to demonstrate this fact. Several features of the cancellation model may be to its advantage: it is closely related to the operation of harmonic cancellation that can account for segregation of concurrent harmonic stimuli, it can be generalized to explain the perception of multiple pitches, and it shows a greater degree of sensitivity to phase than autocorrelation, which may allow it to explain certain phenomena that autocorrelation cannot account for. © 1998 Acoustical Society of America. [S0001-4966(98)00902-3]

PACS numbers: 43.10.Ln, 43.66.Ba, 43.66.Hg, 43.64.Bt [JWH]

## INTRODUCTION

The autocorrelation model of pitch perception dates back to Licklider's "duplex" model of pitch perception (Licklider, 1951, 1956, 1959, 1962). Licklider imagined a network of delay lines and coincidence counters arranged along two axes: frequency (inherited from peripheral filtering) and delay (over the range of periods that can evoke pitch). The network calculated an array of autocorrelation functions (ACF), one for each channel of the peripheral filter. In response to a periodic tone, activity within the network was greatest along a "ridge" spanning the frequency dimension at a lag equal to the period. [An example of a similar pattern produced by the model of Meddis and Hewitt (1991a, b) is displayed in Fig. 3(a).]

Despite its appeal, Licklider's model fell on a blind spot of auditory theory. For many years, favor went mainly to the "pattern matching" models of pitch perception of Wightman (1973), Terhardt (1974), and Goldstein (1973), in spite of the fact that they are tributary to a high-resolution spectral analysis (also present in Licklider's model, but of secondary importance). They also require explicit pattern matching mechanisms: Fourier transformation in the case of Wightman (1973), and learning in the case of Terhardt (1974), whereas pattern matching comes "for free" in Licklider's model. In his model, the fundamental period is calculated simply by looking across channels for a peak common to several channels, and this operation succeeds whether individual components are resolved within channels or not. Pattern matching models are not specific about how pattern matching is implemented physiologically, whereas Licklider's model is both specific and reasonably plausible.

Licklider's model lacked a precise rule to derive a quantitative pitch estimate. This was repaired by Meddis and

Hewitt (1991a, b; Meddis and O'Mard, 1997), who chose perhaps one of the simplest possible rules: the autocorrelation pattern is summed across the frequency dimension to obtain a summary autocorrelation function (SACF) and the pitch is derived from the first major peak in this function (the "period peak") [Fig. 4(a)]. Meddis and Hewitt also undertook the task of simulating the behavior of their model in response to a set of stimuli that evoke pitch phenomena important for pitch theory (missing fundamental, pitch shifts of inharmonic complexes, etc.). This helped dispel several common misconceptions, such as that the ACF might be excessively sensitive to phase if calculated in the time domain, or on the contrary excessively *insensitive* to phase. The hair cell transduction model of Meddis (1988) included in the simulation of their pitch model was instrumental in demonstrating second-order effects such as phase sensitivity that may arise in a physiologically realistic implementation of the autocorrelation model. The "pitch dominance region," of crucial importance for the thesis of pattern matching, emerged in their model as a consequence of the relative weights of low-versus high-frequency channels, and the breakdown of neural firing synchrony at high frequencies. This was corroborated by Cariani and Delgutte (1996a, b), who recorded from the auditory nerve of the cat in response to a range of stimuli important for auditory theory, and showed that in most cases the pitch could be readily derived from the shape of autocorrelation histograms (ACH).

The ACH (or all-order interspike interval histogram) used by Cariani and Delgutte is a relatively recent way of processing recordings from auditory neurons (Ruggero, 1973; Boenger, 1974; Evans, 1983; Shofner, 1991). More common has been the first-order interspike interval histogram (ISIH) (Rose *et al.*, 1967). The abundance of physiological data reported in this format led to several models similar to those of Licklider or Meddis and Hewitt (1991a, b), but using arrays of ISIHs rather than autocorrelation functions (Moore, 1977; van Noorden, 1982).

<sup>a)</sup>"Selected research articles" are ones chosen occasionally by the Editor-in-Chief that are judged (a) to have a subject of wide acoustical interest, and (b) to be written for understanding by broad acoustical readership.

Licklider's model has inspired computational models of pitch (Lyon, 1984; Lazzaro and Mead, 1989; Slaney, 1990) and auditory scene analysis (Lyon, 1983; Weintraub, 1985; Cooke, 1991; Brown, 1992; Meddis and Hewitt, 1992; Lea, 1992) involving two-dimensional autocorrelation arrays. The array is usually summed across the frequency dimension to obtain a summary function similar to the SACF of Meddis and Hewitt. This summary ACF is close in shape to the ACF of the *raw waveform*, and the latter can be used as a simpler model of similar predictive power (Yost, 1996; Yost *et al.*, 1996). The pulse-ribbon and strobed auditory integration (SAI) models of Patterson (Patterson, 1987; Patterson *et al.*, 1992) are related to autocorrelation. The SAI is in effect the cross correlation between the neural response within each channel with a strobe signal consisting of a single pulse per stimulus period, and the pattern it produces is visually quite similar to autocorrelation.

In its modern form(s), the autocorrelation model of pitch appears to becoming accepted to the same degree as "pattern-matching" models. Nevertheless it is not altogether without problems. For one thing, there is as yet little evidence of a two-dimensional autocorrelation map organized according to frequency and lag, despite some evidence for amplitude modulation maps in the inferior colliculus of the cat (Langner and Schreiner, 1988; Langner, 1992). Licklider (1959) noted, however, that an orderly layout should not necessarily be expected. More generally, there is not strong evidence of delay lines of sufficient duration (up to 10–20 ms). This weakness is not entirely specific to autocorrelation. Other models require elements not very different from delay lines (the SAI's signal buffer, for example), or else are rather vague about the physiological operations that they would involve (a "Fourier transformer," for example). It would be unfair to fault autocorrelation for making its requirements explicit. In favor of autocorrelation is evidence for the closely related cross-correlation model of binaural interaction (Jeffress, 1948) found in the MSO and IC of the cat (Yin and Chan, 1990; Yin *et al.*, 1987) and equivalent centers in the owl (Konishi *et al.*, 1988).

Autocorrelation is also difficult to reconcile with some aspects of pitch perception. It works equally well for stimuli consisting of unresolved harmonics as for resolved harmonics, yet there is evidence that the resolved harmonics dominate the pitch percept (Houtsma, 1995). Pitches from resolved and unresolved channels should be readily comparable, yet Carlyon and Shackleton (1994) found that such is not the case. Kaernbach and Demany (1996) found that the pitch of high-pass filtered pulse trains depended on the statistics of first-order intervals between pulses, rather than all-order intervals as one would expect based on autocorrelation.

Another example that the AC model does not readily account for is the difference in percept evoked by certain time-reversed stimuli. Patterson (1994a, b) presented subjects with stimuli consisting of sine-wave carriers shaped with repeated "damped" (exponential decay) or "ramped" envelopes (same shape, but time reversed). Examples are shown in Fig. 9. The carrier frequency is 800 Hz, the repetition rate 40 Hz, and the half-time (time to it takes for the

wave to decrease by half) is a 4 ms. These stimuli will be used later on in this paper for illustration purposes. Both evoked a pitch corresponding to the sine wave carrier, but the pitch was much stronger for ramped than for damped sine waves. The two stimuli have identical spectra and autocorrelation functions, and the AC model does not readily explain why they evoke different sensations.

Despite these problems, the AC model remains a good first-order model, attractive in terms of simplicity, explanatory power, and physiological plausibility. The purpose of this paper is to point out that a similar model can be obtained by replacing the multiplication by subtraction (or excitatory by inhibitory neural interaction). This new model will be referred to loosely as the "cancellation model of pitch perception." The equivalence between autocorrelation and cancellation is illustrated for three variants of the AC model (waveform based, discharge probability based, spike based). On the strength of this equivalence we may assume that major pitch phenomena explainable by autocorrelation models (Meddis and Hewitt, 1991a, b; Cariani and Delgutte, 1996a, b; Yost *et al.*, 1996) are equally well explained by their cancellation counterparts.

Although cancellation and autocorrelation pitch models are equivalent to a first approximation, several features of cancellation make it attractive. For one, cancellation is a flexible and powerful "building block" in that it leaves a residue that may be analyzed in turn, in a succession of estimate-cancel-estimate operations. For another, the way cancellation may be implemented physiologically makes it relatively sensitive to time reversal, and this might account for the time-order-dependent phenomena reported by Patterson (1994a, b).

## I. SUBTRACTION VERSUS MULTIPLICATION AT THE WAVEFORM LEVEL

The autocorrelation function of a waveform  $s(t)$  may be understood as the result of applying a kind of nonlinear "filter" to the waveform

$$P_{\tau}(t) = s(t)s(t - \tau) \quad (1)$$

and then integrating over time

$$\text{ACF}(\tau) = \int_{-\infty}^{\infty} P_{\tau}(\theta) d\theta. \quad (2)$$

In practice the summation is usually performed over a sliding window to obtain a running autocorrelation function indexed by time:

$$\text{ACF}_t(\tau) = \int_{-\infty}^t w(\theta - t) P_{\tau}(\theta) d\theta, \quad (3)$$

where  $w(\theta - t)$  is a window function that emphasizes values near  $t$ . The integration can be seen as a kind of low-pass filter that smooths the quantity  $P_{\tau}(t)$  so that the ACF pattern does not fluctuate too much over time.

As an example, let us consider a waveform made up of ten equal amplitude harmonics of 200 Hz in sine phase (Fig. 1). The ACF of this waveform is shown in Fig. 2(a). The ACF shows peaks at the origin and at 5 ms, period of the

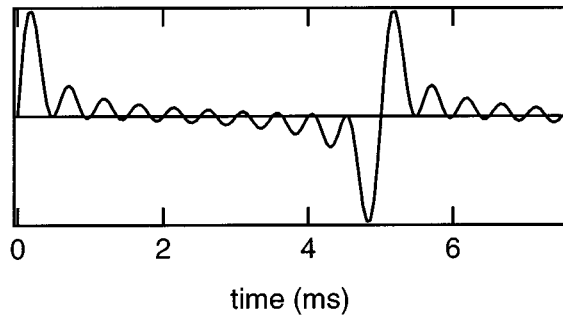


FIG. 1. Waveform consisting of the first ten harmonics of 200 Hz added with equal amplitudes in sine phase.

waveform, as well as at multiples of 5 ms (not visible in the figure). The position of the first “period peak” is the cue to the pitch according to models of pitch perception based on the autocorrelation function of the waveform (Yost, 1996). In this example, summation was performed over a square window covering two periods of the waveform (10 ms). In general, in this paper, integration windows are chosen so as to minimize fluctuation of the pattern (ACF or other) over time, rather than according to psychophysical or physiological estimates. If the windows were chosen too short (for example, shorter than the fundamental period), the shape of the pattern would differ according to where in time it was sampled, with the risk of inconsistent or misleading conclusions. A model using short windows is incomplete unless it specifies how such a “pulsating” pattern is processed to obtain a relatively stable percept such as pitch.

The product of  $s(t)$  and  $s(t-\tau)$  may be replaced by their squared difference  $D_\tau(t)=[s(t)-s(t-\tau)]^2$  to obtain an average “squared difference function” (SDF):

$$\text{SDF}_t(\tau) = \int_{-\infty}^t w(\theta-t) D_\tau(\theta) d\theta. \quad (4)$$

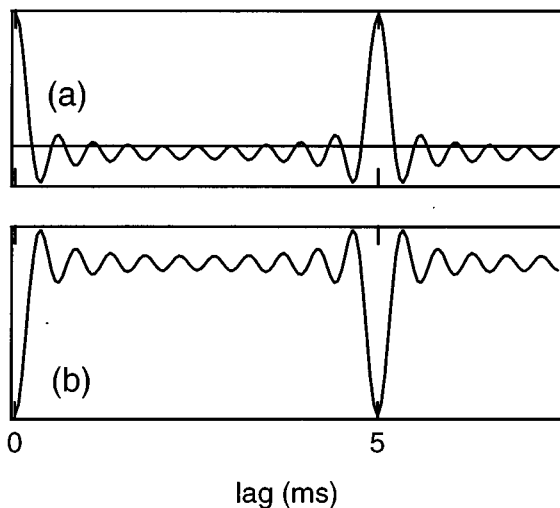


FIG. 2. (a) Running autocorrelation function (ACF) of the waveform of Fig. 1, calculated over a 10-ms window. (b) Average squared difference function (SDF) of the same waveform.

Figure 2(b) shows an example of this function calculated using the same window as for the ACF. The SDF looks like a mirror image of the ACF. Whereas the ACF showed a peak at the origin and at the period (and its multiples), the SDF shows dips in those positions. In fact,  $D_\tau(t)$  and  $P_\tau(t)$  are related:

$$D_\tau(t) = s^2(t) + s^2(t-\tau) - 2s(t)s(t-\tau) \quad (5)$$

$$= P_0(t) + P_0(t-\tau) - 2P_\tau(t) \quad (6)$$

implying a similar relation between SDF and ACF. If the integration window is large enough, fluctuations of  $\text{ACF}_t(\tau)$  with  $t$  are small and we have

$$\text{SDF}_t(\tau) = 2[\text{ACF}_t(0) - \text{ACF}_t(\tau)]. \quad (7)$$

Comparing Fig. 2(a) and (b), it is clear that the SDF could replace the ACF as the basis of a pitch perception model, if the “pitch peak” cue were replaced by a “pitch valley” cue. A pattern similar to the SDF might arise within a coincidence network like that of Licklider, but involving *inhibitory* rather than *excitatory* interaction between time-domain patterns of neural activity. This idea is explored in more detail in the next two sections.

It should be noted that the SDF is closely related to the average magnitude difference function (AMDF) that has been used to estimate the fundamental frequency ( $F_0$ ) of speech (Ross *et al.*, 1974). The difference between the two, apart from the discrete sample notation, is that the AMDF sums absolute values whereas the SDF sums their squares:

$$\text{AMDF}_i(\tau) = \sum_{j=i}^{i+N} |s(j) - s(j-\tau)|. \quad (8)$$

## II. SUBTRACTION VERSUS MULTIPLICATION OF DISCHARGE PROBABILITY

Instead of the waveform, we consider discharge probability within each channel of a model of peripheral filtering and hair-cell transduction (Meddis and Hewitt, 1988). Following the model of Meddis and Hewitt (1991a, b), the ACF of each probability function was calculated to obtain a pattern of activity over two dimensions: characteristic frequency (inherited from peripheral frequency analysis) and the lag dimension of the autocorrelation function. Figure 3(a) shows the pattern evoked by the previous pulse train. A “ridge” is visible at 5 ms, period of the waveform, as well as at the origin. If the ACFs are summed across channels, the resulting summary autocorrelation function (SACF) shows a peak at 5 ms [Fig. 4(a)]. The position of this “period peak” is the cue to pitch in Meddis and Hewitt’s (1991a, b) pitch perception model. In this example the ACF used a 10-ms square integration window rather than the 2.5-ms exponential window used by Meddis and Hewitt.

The ACF pattern might represent the activity across a neural network of delay lines and coincidence neurons in which direct and delayed spike trains interact in *excitatory* fashion. It is possible to imagine a similar network in which the spike trains would interact in *inhibitory* fashion. For example, a coincidence detector might fire with a probability

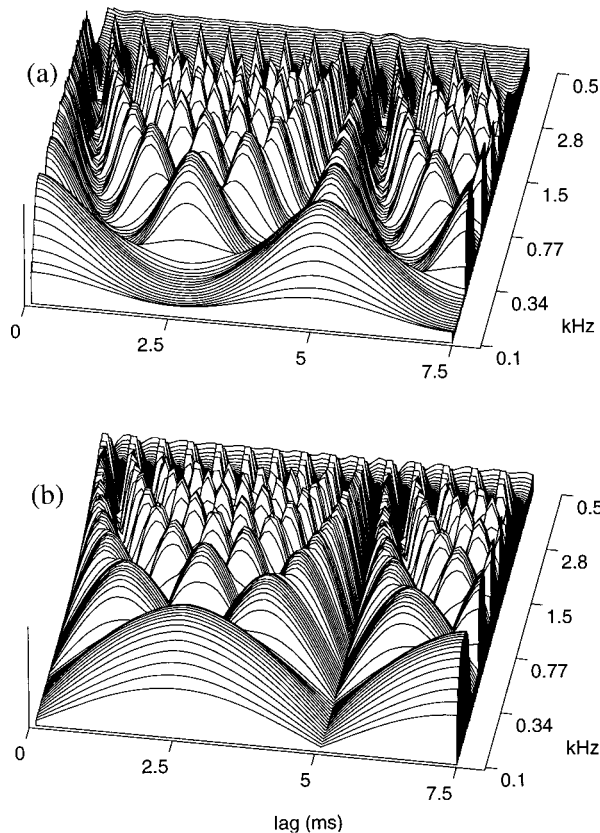


FIG. 3. (a) Array of autocorrelation functions calculated from discharge probabilities produced by a model of peripheral filtering and transduction in response to the waveform of Fig. 1. There are 100 channels, spaced uniformly between 100 and 5000 Hz on an ERB scale. (b) Array of average half-wave rectified difference functions (RDF) in response to the same waveform.

proportional to the difference between spike densities along direct and delayed pathways, as assumed in the “neural cancellation filter” of de Cheveigné (1997a):

$$R_r(t) = \max(0, s(t) - s(t - \tau)), \quad (9)$$

where  $s(t)$  and  $R_r(t)$  are instantaneous firing probability densities at the input and output of the cancellation filter, respectively. The half-wave rectifying  $\max(\cdot)$  operation reflects the fact that probabilities cannot be negative [the firing probabilities  $s(t)$  themselves are approximate versions of the half-wave rectified basilar membrane motion, so half-wave rectification occurs twice within the model]. The activity within an array of such filters, indexed by delay, may be represented by a function similar to the SDF, but in which the squared difference is replaced by a *half-wave rectified* difference. We denote this average “rectified difference function” as RDF:

$$\text{RDF}_i(\tau) = \int_{-\infty}^t w(t - \theta) R_r(\theta) d\theta. \quad (10)$$

Figure 3(b) shows the pattern evoked by the previous pulse train. A “valley” is visible at 5 ms, period of the waveform, as well as at the origin. If the RDFs are summed across channels, the resulting summary half-wave rectified difference function (SRDF) shows a dip at 5 ms [Fig. 4(b)]. The

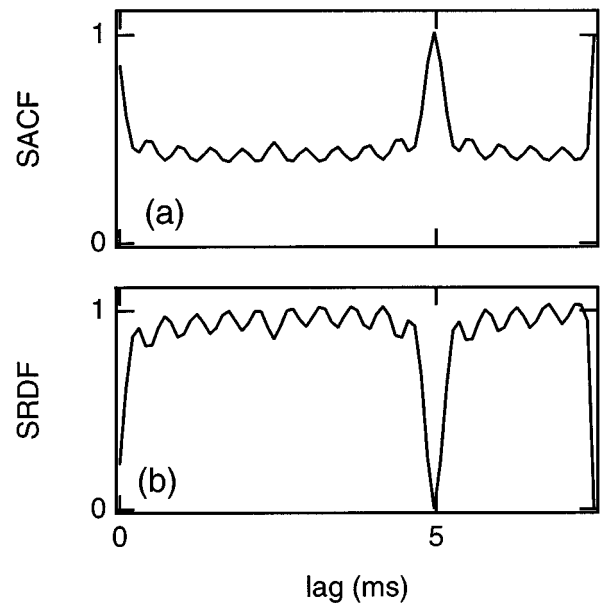


FIG. 4. (a) Summary autocorrelation function (SACF) in response to the waveform of Fig. 1. (b) Summary average half-wave rectified difference functions (SRDF) in response to the same waveform.

position of this “period dip” could be taken as the cue to pitch, resulting in a pitch perception model very similar to that of Meddis and Hewitt (1991a, b). It is interesting to note that, for a perfectly periodic stimulus such as this one, the background-to-dip ratio is infinite for the SRDF, whereas the peak-to-background ratio is finite (about 2 in this case) for the SACF.

Equation (9) represents a hypothetical gating neuron that fires with a probability proportional to the *difference* of firing probabilities at its excitatory and inhibitory synapses. More general interaction may be easier to model explicitly with spike trains, either recorded physiologically or generated by a stochastic spike generation model. That is the subject of the next section.

### III. INHIBITORY VERSUS EXCITATORY COINCIDENCE COUNTING

#### A. Spike generation model

Instead of the raw waveform, or discharge probabilities, we consider spike trains produced by a model of spike generation, driven by probability functions produced by the previous model of peripheral filtering and haircell transduction. Spike times were generated stochastically according to an inhomogeneous Poisson process with a refractory period (Schroeder and Hall, 1974; Johnson, 1980; Johnson and Swami, 1983; Carney, 1993). The model produces lists of “spike” times similar to those recorded in physiological experiments. Histograms (peristimulus, period, interval, autocorrelation) derived from model spike data are similar to those reported in the literature, suggesting that the model reproduces the essential aspects of spike train statistics.

Figure 5(a) shows an example of an autocorrelation histogram (ACH) derived from the 500-Hz channel of the model in response to 100 repetitions of a 100-ms portion of

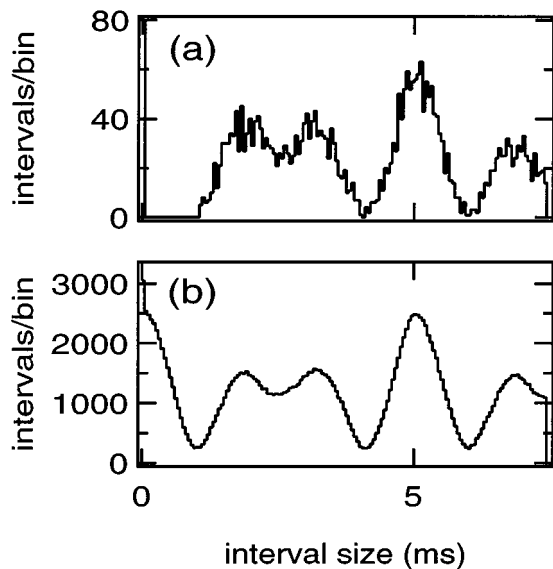


FIG. 5. (a) Autocorrelation histogram (ACH) of a single-fiber spike train produced by the spike generation model in response to the waveform of Fig. 1. The “fiber’s” characteristic frequency was 500 Hz, and the histogram was calculated from 100 repetitions of the 100-ms stimulus. The first 20 ms of each spike train were discarded to remove the onset transient response. Bin width is 100  $\mu$ s. (b) The ACH calculated from the same data, sorted to simulate the activity of 100 fibers of similar characteristics. Note the reduced “noise,” and the lack of the gap due to refractory effects.

the waveform displayed in Fig. 1 (the initial 20 ms of each spike train were discarded to eliminate the transient response at onset). The “gap” below 1 ms is due to the refractory period, and the peak at 5 ms reflects the period of the waveform. This histogram reflects the statistics of intervals between spikes within a *single* fiber. It is reasonable to assume that the auditory system might take into account interspike intervals within a *group* of similar fibers, as this makes more efficient use of the available information. In practice, the interval statistics of a group of  $N$  similar fibers can be simulated by taking the list of spike times for  $N$  presentations of the stimulus to the same fiber (time being measured relative to stimulus onset), and sorting it. Figure 5(b) shows a histogram obtained after sorting spike data in this fashion. The “spikes” are the same as displayed in Fig. 5(a), but the ACH is less noisy and lacks a gap at short intervals. The sorted-spike ACH is equivalent to the autocorrelation of the peristimulus histogram (PST) (Palmer, 1992)

The sample at zero in Fig. 5(b) counts coincidences of spikes “with themselves,” and is equal to the number of spikes in the spike train. This sample is often not represented in histograms of neural data, but its presence is congruent with the definition of autocorrelation. In Fig. 5(a) the value of that bin (3048) is outside the range of the graph. The value of the zero lag bin does not depend on bin width, whereas that of all other bins does. The prominence of this bin relative to the rest of the histogram thus depends on bin width.

## B. Excitatory coincidence network

Licklider’s (1951) model involved an array of coincidence or gating neurons similar to the one schematized in Fig. 6(a). The neuron fires if and only if spikes arrive simul-

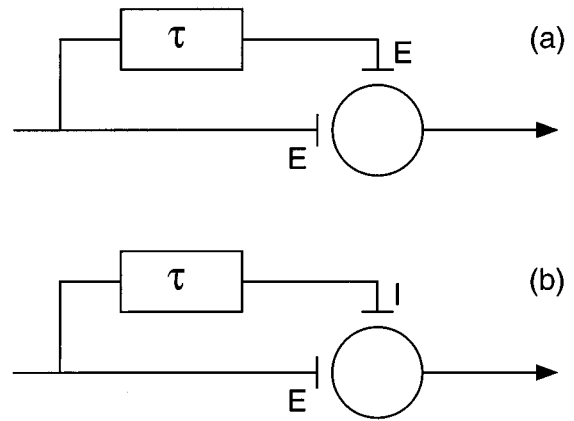


FIG. 6. (a) Excitatory gating neuron. The neuron fires if spikes arrive simultaneously along the direct and delayed pathways. (b) Inhibitory gating neuron. The neuron fires if a spike arrives along the direct path, unless a spike arrives simultaneously along the delayed path.

aneously via direct and indirect pathways (with a certain tolerance). More precisely, each spike arriving along the delayed pathway opens a “window” during which a spike arriving along the direct pathway may be transmitted. In our simulation the window was square and its duration was 0.05 ms (exponential windows were also tested with roughly similar results). Sorted spike trains produced by the spike generator were fed to an array of gating neurons covering a range of delays from 0 to 7.5 ms with 0.02-ms resolution. The number of spikes transmitted as a function of delay is displayed in Fig. 7(a). This plot is similar in aspect to the autocorrelation histogram [Fig. 5(b)]. The size of the coincidence window determines the relative height of the sample at zero lag.

If a similar simulation is performed in other channels

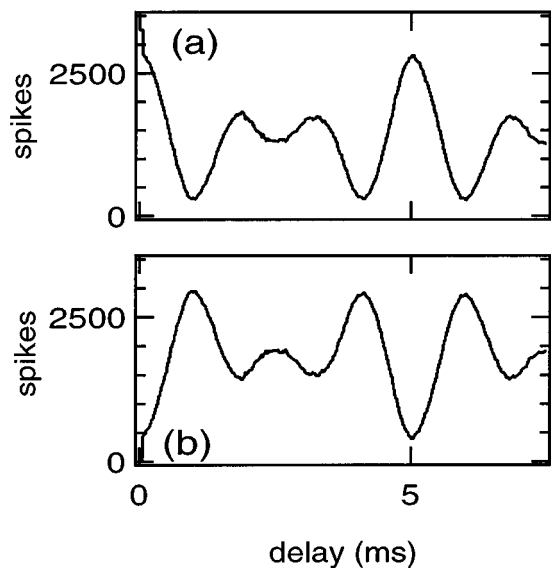


FIG. 7. (a) Pattern of activity at the output of an array of excitatory gating neurons [Fig. 6(a)] in response to the waveform of Fig. 1. The channel’s characteristic frequency is 500 Hz. The coincidence window is square with a width of 0.05 ms. Delays are sampled with a resolution of 0.02 ms. (b) Pattern of activity at the output of an array of inhibitory gating neurons [Fig. 6(a)] in response to the same waveform. Parameters are the same as for (a).

(not shown), a peak appears at the same position (5 ms), as predicted by Licklider and verified by Cariani and Delgutte (1996a, b) with recordings from the auditory nerve.

### C. Inhibitory coincidence network

The excitatory gating neuron of Fig. 6(a) may be replaced by the inhibitory gating neuron of Fig. 6(b). Each spike arriving along the delayed pathway opens a “window” during which a spike arriving along the direct pathway will *not* be transmitted. An array of such inhibitory gating neurons was simulated with the same data and similar parameters as for the excitatory network. The result is plotted in Fig. 7(b). The pattern is the mirror image of that plotted in Fig. 7(a). The dip at 5 ms can serve as a cue to the pitch of the stimulus, playing the same role as the peak in Fig. 7(a). The size of the coincidence window determines the depth of this dip.

In summary, in each version of the AC model, multiplication can be replaced by subtraction (or excitatory gating by inhibitory gating) to produce an equivalent cancellation model. To a first approximation the behavior is the same, and these cancellation-based models can account for the same class of pitch effects as their autocorrelation-based counterparts. In addition to the examples reported here, the models were simulated with a variety of stimuli (missing fundamental, inharmonic complexes, synthetic vowels, rippled noise, etc.). Peaks and dips generally occurred at the same positions for autocorrelation and cancellation, although in some cases there were differences in the strength of the pitch cue (Sec. IV B).

## IV. WHY SUBTRACT?

So far we have emphasized the similarity between models based on multiplication and subtraction. Why then consider the latter, given that autocorrelation models are successful and well established? To the extent that the two are equivalent, a cancellation-based model may be regarded as an alternative implementation of its autocorrelation-based counterpart. Even if it performs no better than autocorrelation, cancellation offers a fresh perspective. Physiologists might find use for a model in which *minima* of activity are given the importance usually attributed to *maxima*. In addition, cancellation is a flexible “building block” for auditory modelling, and its implementation implies properties that might explain second-order effects that AC models have difficulty accounting for.

### A. Cancellation as a building block of auditory processing

There is evidence, from experiments on concurrent vowel identification, that concurrent harmonic sounds are segregated by a mechanism of *cancellation* of harmonic interference. Although other models of harmonic cancellation have been proposed (Meddis and Hewitt, 1992), one effective way to perform the operation is with a “neural” filter similar to that illustrated in Fig. 6 or defined by Eq. (9) (de Cheveigné, 1993, 1997a). Harmonic cancellation requires an estimate of the period of the interference. While this can be

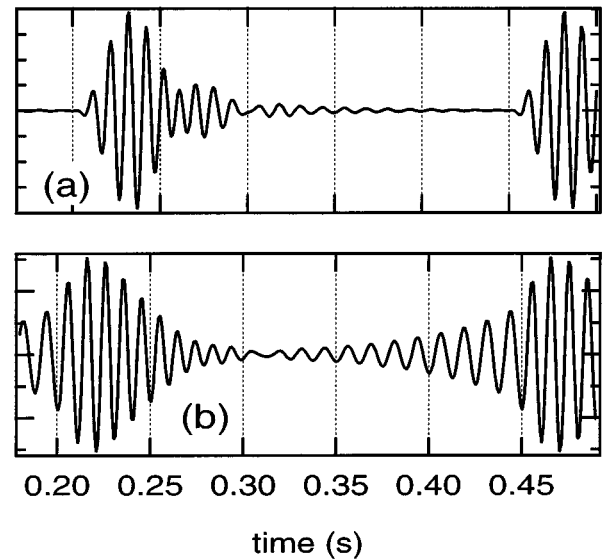


FIG. 8. (a) Damped sine wave stimulus filtered by a gammatone filter of center frequency 1040 Hz. (b) Same for the ramped sine wave.

obtained from a variety of period-estimation models, for example ACF based, it is expedient to derive it from the cancellation filter itself, by searching for a minimum of Eq. (10) as a function of its parameter  $\tau$  (de Cheveigné, 1997a). Given that the period-estimation principle is successful in that context, it makes sense to apply it also to pitch perception.

Harmonic cancellation may also be used to build a model of the perception of *multiple pitches* evoked by concurrent periodic sounds. The situation is common in music when several instruments play together, and trained listeners can accurately estimate the pitches of concurrent periodic sounds, even if there are no differences in onset or spectral envelope (Nordmark, 1978). Multiple periods may be estimated recursively in a succession of estimate–cancel–estimate steps, or in parallel according to a joint cancellation algorithm. The principle was applied with success to  $F_0$  estimation of pairs of natural spoken voices by de Cheveigné (1993), who found that it was superior to several other two-period estimation algorithms. It can be generalized to an arbitrary number of concurrent periodic signals (de Cheveigné and Kawahara, 1997). Examples of the effectiveness of harmonic cancellation for reinforcing the representation of a nondominant period may be found in de Cheveigné (1993, 1997a). If cancellation-based estimation can successfully handle multiple pitches, it makes sense to invoke it in the limit case of a single pitch. A link between harmonic sound segregation and pitch (though not necessarily via this model) was an ingredient of a model that explained the pitch shifts of mistuned partials observed by Hartmann and Doty (1996) (de Cheveigné, 1997b).

Harmonic cancellation allows interfering components to be “peeled away” from a target, but it can also serve another purpose. Responses of single peripheral channels to stimuli with transient waveforms show ringing effects due to basilar membrane filtering. This is illustrated in Fig. 8 for a 1040-Hz channel responding to the damped and ramped stimuli that we mentioned before (Fig. 9). This corresponds to a channel

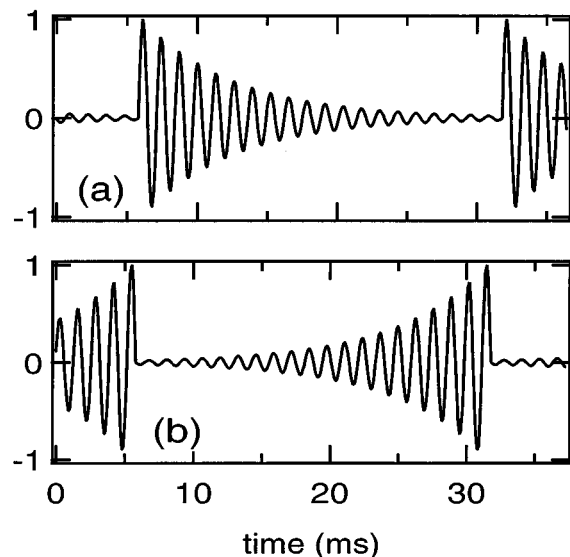


FIG. 9. (a) “Damped” sine wave stimulus as used by Patterson (1994a, b). Carrier frequency is 800 Hz, modulation frequency is 40 Hz, and half-life is 4 ms. (b) Similar “ramped” sine wave stimulus.

tuned slightly higher than the 800-Hz carrier (cf. Figs 6 and 7 of Patterson, 1994a). The response to the ramped stimulus [Fig. 8(b)] can be decomposed into two intervals: an interval of ringing excited by the sharp offset of the previous ramp, with a periodicity of about 1040 Hz, and an interval of gradual onset of the current ramp, with a periodicity of 800 Hz. The response to the damped stimulus is more complex: ringing and stimulus interact throughout the modulation period. The auditory system faces a difficult problem in analyzing such patterns in the time domain, as they reflect both the stimulus and less interesting resonance characteristics.

Certain models incorporate schemes to remove the effects of ringing. Weintraub (1985) subtracted from the 2-D autocorrelation pattern the same pattern obtained in response to white noise. The response to noise reflects only basilar membrane filtering, and subtracting it was expected to remove some effects of filter ringing. The Patterson–Holdsworth software suite (Patterson *et al.*, 1992) includes a temporal adaptation stage designed to reduce the effects of ringing (Holdsworth, 1990). Additionally in that model, as in multichannel autocorrelation models (Meddis and Hewitt, 1991a, b), ringing effects within individual channels are averaged out by summation over channels.

Ringing effects can also be removed by harmonic cancellation. Figure 10(a) represents the basilar membrane filter response of Fig. 8(a) filtered by a cancellation filter tuned to 0.962 ms (inverse of the gammatone filter’s center frequency). Figure 10(b) shows the same processing in the case of the ramped stimulus. Suppression of ringing effects is evident in both cases. The fine structure of the response reflects the stimulus carrier (800 Hz) and its envelope resembles that of the stimulus. There is no evidence of the filter’s ringing periodicity. It is perhaps premature to include a “ringing suppressor” in every model of auditory processing, but this example illustrates well the power and flexibility

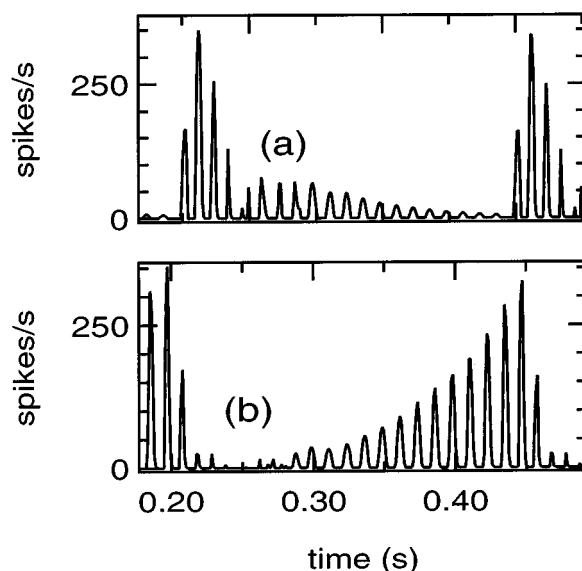


FIG. 10. (a) Damped sine wave stimulus filtered by a gamma tone filter of center frequency 1040 Hz, followed by a hair cell model, followed by a cancellation filter tuned to 0.962 ms (period of the gammatone filter center frequency). (b) Same for the ramped sine wave.

of cancellation as a building block for neural “signal processing” models.

It is interesting to note the following modification that may be of use in the case of an amplitude-varying stimulus or impulse response:

$$R'_\tau(t) = \max(0, s(t) - \alpha s(t - \tau)). \quad (11)$$

A value of  $\alpha < 1$  is sufficient to suppress a decreasing signal, whereas a value greater than 1 is required to suppress a ramped signal. The extra parameter compromises parsimony, but the auditory system (supposing it employs this form of time-domain cancellation) might find such fine tuning useful.

Cancellation has been proposed as a mechanism to explain binaural release from masking (Durlach’s Equalization-Cancellation model, 1963).

## B. Temporal asymmetry

The product  $P_\tau(t) = s(t)s(t - \tau)$  summed in the ACF is symmetric in  $s(t)$  and  $s(t - \tau)$  [Eq. (3)]. The ACF is therefore not affected by a reversal of the time axis. The same is true of the SDF that sums the squared difference [Eq. (4)]. Therefore, neither can account for the greater pitch strength of ramped versus damped stimuli observed by Patterson (1994a, b). The hair cell transduction stage used by Meddis and Hewitt (1991a, b) introduces a certain degree of sensitivity to temporal asymmetry, but Irino and Patterson (1996) showed that it is insufficient to account for the asymmetry in pitch strength. Based on the relative insensitivity of the ACF to time reversal, Patterson (1994b) argued in favor of the strobed temporal integration (STI) process incorporated in his auditory image model (AIM) (Patterson *et al.*, 1995). However, Irino and Patterson (1996) showed that a sensitivity to temporal asymmetry may be the consequence of a variety of other “delta-gamma” processes that arise at various stages of auditory processing. The delta-gamma operator em-

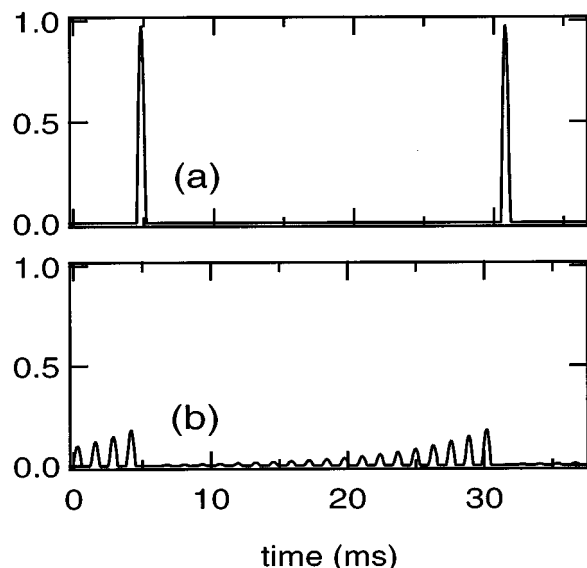


FIG. 11. (a) Cancellation filter output for an input consisting of a half-wave-rectified damped sine wave. Lag is 1.25 ms (period of the 800-Hz carrier). (b) Same for the ramped sine wave. The vertical scale of both graphs was normalized by dividing by the peak output observed in the damped case.

phasizes the portion of signals for which the derivative of the envelope is positive rather than negative. It turns out that the half-wave rectified difference  $R_r(t)$  involved in the RDF model has such a property.

The effect of a time reversal is illustrated in the case of the ramped and damped stimuli in Fig. 11, for a delay of 1.25 ms, period of the 800-Hz carrier. For this illustration, the quantity  $R_r(t)$  was calculated directly from the half-wave rectified stimulus waveform, to avoid compounding effects of basilar membrane filtering and haircell transduction. For the damp, the first peak of each resonance is reflected in  $R_r(t)$ , but each successive peak falls in the “shadow” of its predecessor and is suppressed. The result is a single pulse per modulation period. For the ramped signal (b),  $R_r(t)$  is a ramped series of pulses at the period of the carrier. When peripheral filtering is included the picture is less clear, but it remains true that  $R_r(t)$  is asymmetric and sensitive to time reversal. The consequences of this asymmetry depend upon other details of the model. It turns out the RDF itself, as defined in Eq. (10) is *not* sensitive to a time reversal, despite the strong asymmetry of the quantity  $R_r(t)$  that it integrates.

There are at least three ways by which the asymmetry can be reintroduced. One is to replace  $R_r(t)$  by its square before summation. This is illustrated in Fig. 12(a). The dips in the RDF are deeper for the ramped stimulus (full line) than for the damped stimulus (dotted line). If pitch strength is a function of the depth of the pitch cue, the pitch should be stronger for ramped than for damped stimuli, as observed by Patterson (1994a, b). Note that if a compressive nonlinearity were applied before integration instead of an expansive nonlinearity (square), the opposite asymmetry would have resulted.

A second way to reintroduce an asymmetry is to replace  $R_r(t)$  by  $R'_r(t)$ , as defined in Eq. (11), with a value of  $\alpha$

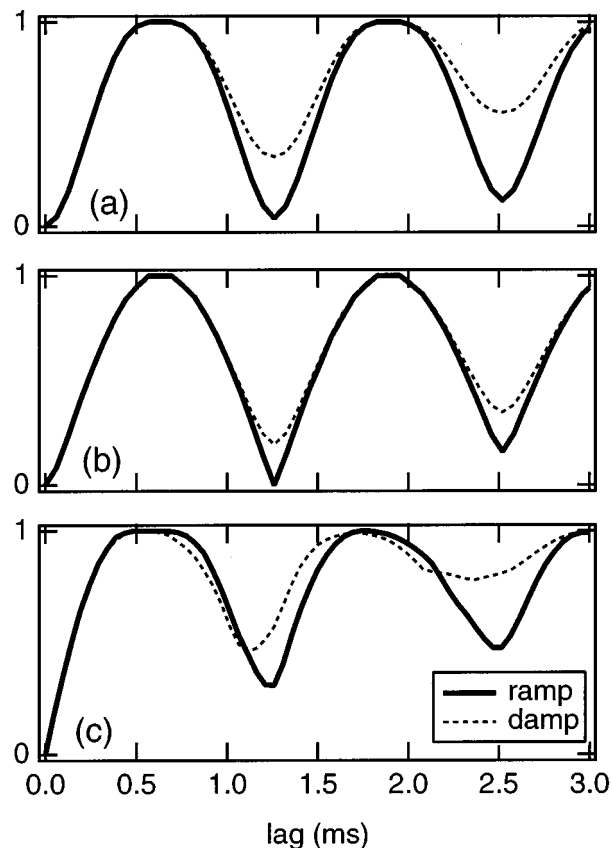


FIG. 12. (a) Modified RDF for a ramped (full line) or damped (dotted line) sine wave. The RDF was modified by raising  $R_r(t)$  to the square before integration. It was calculated from the half-wave rectified waveform (no filtering or haircell transduction). (b) Same, but the RDF was modified by replacing  $R_r(t)$  by  $R'_r(t)$  [Eq. (11)] with  $\alpha = 1.3$ . (c) The RDF calculated from a single channel of the basilar membrane model tuned to 1040 Hz with ringing suppression, in response to the ramped (full line) and damped (dotted line) stimuli. The channel was processed by half-wave rectification followed by a cancellation filter tuned to 1040 Hz (ringing suppression), followed by the unmodified RDF.

greater than 1. This is illustrated in Fig. 12(b). The dips in the RDF are considerably deeper for ramped than for damped stimuli. Again, the opposite asymmetry would have resulted if  $\alpha < 1$ . A third way to introduce temporal asymmetry is to assume ringing suppression as described in Sec. IV A. As evident in Fig. 10, the ramped waveform is more successfully salvaged than the damped waveform. This results again in deeper dips in the RDF for ramped than for damped stimuli [Fig. 12(c)].

In order to determine whether such asymmetries might be sufficient to account for the experimental results of Irino and Patterson (1996), ramped stimuli of half-lives 4, 8, 16, and 32 ms were matched by damped stimuli with half-lives chosen to produce the same depth of the period cue, according to the procedure of Irino and Patterson. An asymmetry factor was calculated using their equation. The first scheme [Fig. 12(a)] yielded a value of 2.16, close to the value of 2.3 observed experimentally for sinusoidal carriers. For the second scheme [Fig. 12(b)], it was impossible to find a damped stimulus long enough to match the extremely deep period dip obtained for ramps. There is no upper limit to the asymmetry



factors that can be “predicted” with this scheme. Finally, the third scheme [Fig. 12(c)], based on a single 1040-Hz channel, yielded an asymmetry factor of 0.95, and other channels yielded similar values. The values of these three schemes bracket those observed experimentally.

In conclusion, the “neural cancellation filter” is inherently sensitive to a time reversal, and this can readily result in a difference in the strength of pitch cues. The ACF is by definition insensitive to phase, although hair-cell adaptation and transduction introduce a certain degree of phase sensitivity in practice. A short integration window also causes phase dependencies, but a model based on a short window is not properly defined unless it specifies how the fluctuating pattern (ACF or other) should be sampled in time. Differences in sensitivity to phase are a minor exception to our main conclusion that ACF and cancellation models are equivalent and can be used interchangeably.

## V. SUMMARY

Multiplication or excitatory neural interaction as used in the family of autocorrelation models of pitch perception can be replaced by subtraction or inhibitory neural interaction to form a corresponding family of cancellation models. The behavior of the models is the same to a first approximation, and cancellation models can account for much the same set of pitch phenomena as autocorrelation models. To a second approximation, the cancellation models differ in some ways from their autocorrelation counterparts. Whereas autocorrelation is largely phase insensitive, cancellation-based models show a sensitivity to phase that may allow them to account for phenomena such as the different sensations evoked by stimuli reversed in time. Cancellation is more flexible as a “building block” for auditory processing than multiplicative or excitatory interaction, as it leaves a “residue” that can be analyzed in turn. Cancellation models can thus be built to account for the perception of multiple pitches evoked by concurrent harmonic sounds such as simultaneous musical notes. Harmonic cancellation seems to be a major mechanism underlying  $F_0$ -guided segregation, and it makes sense to assume that the other main “client” of periodicity, namely pitch, is derived from a cancellation-based mechanism. Indeed, given that the need for sound segregation is probably older and more important for survival than musical pitch perception, one could speculate that pitch might be a “spin-off” of mechanisms that evolved for the purpose of sound segregation.

The cancellation model has the same requirement of delay lines as does the AC model (and also neural cancellation models of  $F_0$ -guided segregation), and the lack of physiological evidence for these delay lines is a major problem. It is conceivable that the delay lines exist but that technical difficulties prevent recording from them directly. It is also possible that patterns involving *minima* of activity were overlooked by researchers expecting *maxima*. If so, the different perspective taken by the present model might be of use in future investigations.

What sets the AC model apart from the pattern matching models of Wightman (1973), Terhardt (1974), and Goldstein (1973) is not so much its principle of time-domain periodic-

ity estimation as the fact that it does not require a separate stage of pattern matching. For “unresolved” channels the pitch cue is obtained directly from the ACF and from “resolved” channels it is obtained by simple comparison or summation of ACFs across channels. Pattern matching is often thought of as being based on place cues, but this is not essential. Goldstein and Srulovicz (1977) suggested that the spectral pitches required for pattern matching could be determined from auditory nerve fiber interspike interval statistics, measured by ISI histograms. The ISI histogram can be replaced advantageously in this role by the AC histogram, and autocorrelation in turn might be replaced by cancellation, to form a pattern-matching model in which spectral pitch cues are cancellation based. The advantage of cancellation over the AC or ISI histograms in this context is that partials that are too close to be resolved by peripheral filtering might be resolved within channels by the multiple period estimation version of the cancellation model. That is important if one is to account for the perception of the pitches of concurrent sounds, whose partials may be very close in frequency. The disadvantage of this proposition, relative to the AC model or the models discussed in Secs. I–III, is of course that it leaves out entirely the issue of how pattern matching is performed. On the other hand, as periods of most harmonics are shorter than the fundamental period, the problem of delay lines is somewhat eased.

Carlyon (1996, 1997) found that two concurrent sounds restricted to a spectral region where their components were not resolvable (3900–5400 Hz) did not evoke two pitches. This contradicts, at least in those conditions, our proposal that several periods may be determined within the same channel. Carlyon has investigated other situations in which perceptual acuity differs between stimuli made of “unresolved” components (high frequency and/or closely spaced) and those made of “resolved” components (low frequency and/or widely spaced). Performance on pitch tasks is usually best for the latter, and comparison between the two may be poor (Carlyon and Shackleton, 1994). All this argues against a single mechanism that treats all spectral regions alike, such as the AC model or the models of Secs. I–III, and rather in favor of a pattern-matching mechanism for the resolved region (for how else can we explain the importance of resolution?). By “resolved,” it is usually understood that peripheral filters are sufficiently sharp to resolve partials of each sound *by itself*. It is not clear how this property might extend to partials of sounds that are mixed, in particular with small  $F_0$  differences ( $\Delta F_0$ ). In one concurrent vowel experiment, a  $\Delta F_0$  of 0.4% improved identification of a vowel that was 15 dB weaker than its competitor. In another, a  $\Delta F_0$  of 3% improved identification of a vowel that was 25 dB weaker than its competitor (de Cheveigné, 1997c). In such difficult conditions, it is hard to imagine a partial of the weaker vowel being isolated within any channel, and this casts doubt on whether “resolvable” partials are actually resolved when  $F_0$ -guided segregation occurs. Within the context of the neural cancellation model that Sec. III is based upon, the advantage of the region of “resolved partials” might be explained by the fact that neural cancellation is imperfect, and better performed after a first step of linear analysis has improved

the signal-to-noise ratio (see also Meddis and O'Mard, 1997 for another explanation). It is clear that this issue requires more investigation.

## ACKNOWLEDGMENTS

Simulation of gammatone filtering and hair cell transduction was based on Malcolm Slaney's "Auditory Toolbox" Matlab code (Slaney, 1993). Minoru Tszuzaki, Peter Cariani, John Culling, Bob Carlyon, Bill Yost, and Ray Meddis made useful comments on earlier versions of the manuscript.

Boerger, G. (1974). "Coding of repetition pitch in the cochlear nucleus of the cat," in *Facts and Models in Hearing*, edited by E. Zwicker and E. Terhardt (Springer-Verlag, Berlin), pp. 206–215.

Brown, G. J. (1992). "Computational auditory scene analysis: a representational approach," Sheffield, Department of Computer Science, unpublished doctoral dissertation.

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. (1996). "Encoding the fundamental frequency of a complex tone in the presence of a spectrally overlapping masker," *J. Acoust. Soc. Am.* **99**, 517–524.

Carlyon, R. (1997). "The effects of two temporal cues on pitch judgments," *J. Acoust. Soc. Am.* **102**, 1097–1105.

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. H. (1993). "A model for the responses of low-frequency auditory-nerve fibers in cat," *J. Acoust. Soc. Am.* **93**, 401–417.

Cooke, M. P. (1991). "Modelling auditory processing and organisation," Sheffield, Department of Computer Science, unpublished doctoral dissertation.

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

de Cheveigné, A. (1997b). "Harmonic fusion and pitch shifts of inharmonic partials," *J. Acoust. Soc. Am.* **102**, 1083–1087.

de Cheveigné, A. (1997c). "Ten experiments in concurrent vowel segregation," ATR Human Information Processing Research Labs, Tech. Rep. TR-H-217.

de Cheveigné, A., and Kawahara, H. (1997). "Modeling the perception of multiple pitches," Proc. IJCAI Workshop on Computational Auditory Scene Analysis, Nagoya (August 1997) (a preliminary draft is available at: <http://www.hip.atr.co.jp/~alain/ijcai97.ps>).

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

Evans, E. F. (1983). "Pitch and cochlear nerve fibre temporal discharge patterns," in *Hearing—Physiological Bases and Psychophysics*, edited by R. Klinke and R. Hartmann (Springer-Verlag, Berlin), pp. 140–146.

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.

Goldstein, J. L., and Srulovicz, P. (1977). "Auditory-nerve spike intervals as an adequate basis for aural frequency measurement," in *Psychophysics and Physiology of Hearing*, edited by E. F. Evans and J. P. Wilson (Academic, London), pp. 337–347.

Hartmann, W. M., and Doty, S. L. (1996). "On the pitches of the components of a complex tone," *J. Acoust. Soc. Am.* **99**, 567–578.

Holdsworth, J. (1990). "Two dimensional adaptive thresholding," APU AAM-HAP Report, Vol. 1, Annex 4, MRC APU, Cambridge.

Houtsma, A. J. M. (1995). "Pitch perception," in *Hearing*, edited by B. C. J. Moore (Academic, London), pp. 267–295.

Irino, T., and Patterson, R. (1996). "Temporal asymmetry in the auditory system," *J. Acoust. Soc. Am.* **99**, 2316–2331.

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

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.

Johnson, D. H., and Swami, A. (1983). "The transmission of signals by auditory-nerve fiber discharge patterns," *J. Acoust. Soc. Am.* **74**, 493–501.

Kaernbach, C., and Demany, L. (1996). "Psychophysical evidence against the autocorrelation theory of pitch perception" (submitted).

Konishi, M., Takahashi, T. T., Wagner, H., Sullivan, W. E., and Carr, C. E. (1988). "Neurophysiological and anatomical substrates of sound localization in the owl," in *Auditory Function—Neurobiological Bases of Hearing*, edited by G. M. Edelman, W. E. Gall, and W. M. Cowan (Wiley, New York), pp. 721–745.

Langner, G. (1992). "Periodicity coding in the auditory system," *Hearing Res.* **60**, 115–142.

Langner, G., and Schreiner, C. E. (1988). "Periodicity coding in the inferior colliculus of the cat. I. Neuronal mechanisms," *J. Neurophysiol.* **60**, 1799–1822.

Lazzaro, J., and Mead, C. (1989). "Silicon modeling of pitch perception," *Proc. Natl. Acad. Sci. USA* **86**, 9597–9601.

Lea, A. (1992). "Auditory models of vowel perception," Nottingham University, unpublished doctoral dissertation.

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

Licklider, J. C. R. (1956). "Auditory frequency analysis," *Information Theory*, edited by C. Cherry (Butterworth, London), pp. 253–268.

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

Licklider, J. C. R. (1962). "Periodicity pitch and related auditory process models," *Int. Audiol.* **1**, 11–36.

Lyon, R. (1984). "Computational models of neural auditory processing," *IEEE ICASSP*, 36.1(1–4).

Meddis, R. (1988). "Simulation of auditory-neural transduction: further studies," *J. Acoust. Soc. Am.* **83**, 1056–1063.

Meddis, R., and Hewitt, M. J. (1991a). "Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I: Phase 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 Hewitt, M. J. (1992). "Modeling the identification of concurrent vowels with different fundamental frequencies," *J. Acoust. Soc. Am.* **91**, 233–245.

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

Moore, B. C. J. (1977). *An Introduction to the Psychology of Hearing* (Academic, London) (reprinted in 1996).

Nordmark, J. O. (1978). "Frequency and periodicity analysis," in *Handbook of Perception, Vol. IV—Hearing*, edited by E. C. Carterette and P. P. Friedman (Academic, New York), pp. 243–282.

Palmer, A. R. (1992). "Segregation of the responses to paired vowels in the auditory nerve of the guinea-pig using autocorrelation," in *Audition Speech and Language*, edited by M. E. H. Schouten (Mouton-DeGruyter, Berlin), pp. 115–124.

Patterson, R. D. (1987). "A pulse ribbon model of peripheral auditory processing," in *Auditory Processing of Complex Sounds*, edited by W. A. Yost and C. S. Watson (Erlbaum, Hillsdale, NJ), pp. 167–179.

Patterson, R. D. (1994a). "The sound of a sinusoid: time-domain models," *J. Acoust. Soc. Am.* **94**, 1419–1428.

Patterson, R. D. (1994b). "The sound of a sinusoid: spectral models," *J. Acoust. Soc. Am.* **96**, 1409–1418.

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

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

Rose, J. E., Brugge, J. F., Anderson, D. J., and Hind, J. E. (1967). "Phase-

- locked response to low-frequency tones in single auditory nerve fibers of the squirrel monkey," *J. Neurophysiol.* **30**, 769–793.
- Ross, M. J., Shaffer, H. L., Cohen, A., Freudberg, R., and Manley, H. J. (1974). "Average magnitude difference function pitch extractor," *IEEE Trans. Acoust., Speech, Signal Process.* **ASSP-22**, 353–362.
- Ruggero, M. A. (1973). "Response to noise of auditory nerve fibers in the squirrel monkey," *J. Neurophysiol.* **36**, 569–587.
- Schroeder, M. R., and Hall, J. L. (1974). "Model for the mechanical to neural transduction in the auditory receptor," *J. Acoust. Soc. Am.* **55**, 1055–1060.
- Shofner, W. P. (1991). "Temporal representation of rippled noise in the anteroventral cochlear nucleus of the chinchilla," *J. Acoust. Soc. Am.* **90**, 2450–2466.
- Slaney, M. (1990). "A perceptual pitch detector," *ICASSP-90*, pp. 357–360.
- Slaney, M. (1993). "An efficient implementation of the Patterson-Holdsworth auditory filter bank," Apple Computer technical report #35.
- Terhardt, E. (1974). "Pitch, consonance and harmony," *J. Acoust. Soc. Am.* **55**, 1061–1069.
- van Noorden, L. (1982). "Two channel pitch perception," in *Music, Mind, and Brain*, edited by M. Clynes (Plenum, London), pp. 251–269.
- Weintraub, M. (1985). "A theory and computational model of auditory monaural sound separation," University of Stanford, unpublished doctoral dissertation.
- Wightman, F. L. (1973). "The pattern-transformation model of pitch," *J. Acoust. Soc. Am.* **54**, 407–416.
- Yin, T. C. T., and Chan, J. C. K. (1990). "Interaural time sensitivity in media superior olive of cat," *J. Neurophysiol.* **64**, 465–488.
- Yin, T. C. T., Chan, J. C. K., and Carney, L. H. (1987). "Effects of interaural time delays of noise stimuli on low-frequency cells in the cat's inferior colliculus. III. Evidence for cross-correlation," *J. Neurophysiol.* **58**, 562–583.
- Yost, W. A. (1996). "Pitch strength of iterated rippled noise," *J. Acoust. Soc. Am.* **100**, 3329–3335.
- Yost, W. A., Patterson, R., and Scheft, S. (1996). "A time domain description for the pitch strength of iterated rippled noise," *J. Acoust. Soc. Am.* **99**, 1066–1078.