Pitch shifts of mistuned partials: A time-domain model

Alain de Cheveigné
Laboratoire de Linguistique Formelle, CNRS/Université Paris 7, 2 place Jussieu, 75251, Paris, France
and ATR Human Information Processing Research Laboratories, 2-2 Hikaridai, Seika-cho,
Soraku-gun, Kyoto 619-0288, Japan

(Received 9 December 1998; revised 17 February 1999; accepted 20 April 1999)

Mistuning one partial of a complex harmonic tone makes that partial easier to hear as a tone separate from the complex. At the same time, two pitch shifts may be observed. First, the low pitch of the complex is shifted in the direction of the mistuning, as if it were ''pulled'' by the partial. Second, the mistuning of the partial is perceptually exaggerated, as if the pitch of the partial were ''pushed'' away from the harmonic series defined by the complex. This paper shows how the latter effect can emerge within a hypothetical neural circuit. The circuit involves a gating neuron fed by three pathways, one direct and excitatory and the other two delayed and inhibitory. The neuron responds to any excitatory input spike unless it is accompanied by an inhibitory input spike on either delayed input, thus acting as a kind of ''anticoincidence counter.'' The first delay is fixed and tuned to the period of the background harmonic complex. Its purpose is to weaken correlates of in-tune components and allow the mistuned partial to stand out. The second delay is variable and used to estimate the period of the mistuned partial, by searching for a minimum output as a function of delay. With an appropriate choice of parameters, the estimate is subject to shifts that are of the same sign as the mistuning and that peak at about 4% mistuning and decrease beyond, as observed experimentally. © *1999 Acoustical Society of America.* [S0001-4966(99)02208-0]

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

INTRODUCTION

Partials of a harmonic complex are difficult to hear unless attention is drawn to them, for example, with the help of a resonator (Helmholtz, 1877). However, if a partial is mistuned it becomes easier to hear as an entity separate from the complex (Hartmann *et al.*, 1986, 1990; Moore *et al.*, 1986). At the same time, both the pitch of the partial and that of the complex are affected by the mistuning. The pitch of the partial varies with mistuning, but more than expected on the basis of its frequency, as illustrated in Fig. 1(a). The difference between measured pitch (symbols) and expected pitch (dotted line) is known as the ''pitch shift of the mistuned partial'' (Hartmann and Doty, 1996). The shift has typically the same sign as the mistuning; it is greatest when the mistuning is about 4%, and it decreases for larger mistunings [Fig. 1(b)].

The low pitch of the complex also changes with the mistuning of the partial. The shift is in the same direction as the mistuning and peaks at a mistuning of about 3% [Fig. 1(c)], beyond which the low pitch starts to move back until at 8% mistuning it regains its original value (Moore *et al.*, 1986; Darwin *et al.*, 1994; Hukin and Darwin, 1995). When plotted, both shifts look alike but their nature is rather different: The low pitch is as it were ''attracted'' by the mistuned partial, whereas the pitch of the partial is ''repulsed'' by the harmonic series. The mutual influence between partial and complex is not symmetrical.

The existence of a low pitch shift is relatively easy to account for. Most pitch models base their estimate of the low pitch on an ensemble of partials, and would respond to a small mistuning of any of them by a proportional shift of the low pitch (Terhardt, 1974; Goldstein, 1973; Meddis and O'Mard, 1997). In percentage, the slope of the shift-vs-mistuning function for any partial is less than one [compare with the dotted line in Fig. 1(c)], and its value can be used as an estimate of the weight of the partial in the pitch calculation (Moore *et al.*, 1985; Moore, 1987). An extra assumption is required to explain why this dependency is limited to mistunings smaller than 3%, beyond which the weight of the partial decreases. For that purpose, Duifhuis *et al.* (1982) proposed a ''harmonic sieve'' to limit the amount of mistuning permitted to partials participating in the pitch. The ''sieve'' may take the form of an arbitrary rule in a pattern matching algorithm, or else it may be attributed to an emergent property of a pitch extraction mechanism. For example, in a model based on histograms of first-order interspike intervals (ISI) pooled across frequency, ISIs evoked by the mistuned partial affected the position of the ''period mode'' at small mistunings, and formed a separate distribution at large mistunings (Moore, 1987).

In contrast, the pitch shift of the mistuned partial is harder to explain. Most pitch models focus on the pitch of the whole stimulus (low pitch) and do not explicitly address the perception of the pitch of individual partials. (Licklider, 1951; Moore, 1977; Goldstein, 1973; Wightman, 1973; Meddis and Hewitt, 1991a, b). Srulovicz and Goldstein (1983) mentioned that neighboring partials might affect the *accuracy* of the representation of a partial's frequency, but said nothing of effects upon its value. Goldstein *et al.* (1978) invoked small biases of the estimates of components of a complex, but did not say how they arose, or how to predict their value. Terhardt (1979) gave precise rules determining how the pitch of a partial (''spectral pitch'') should be affected by the presence of other partials: The fundamental component should shift negatively, and all other components positively.

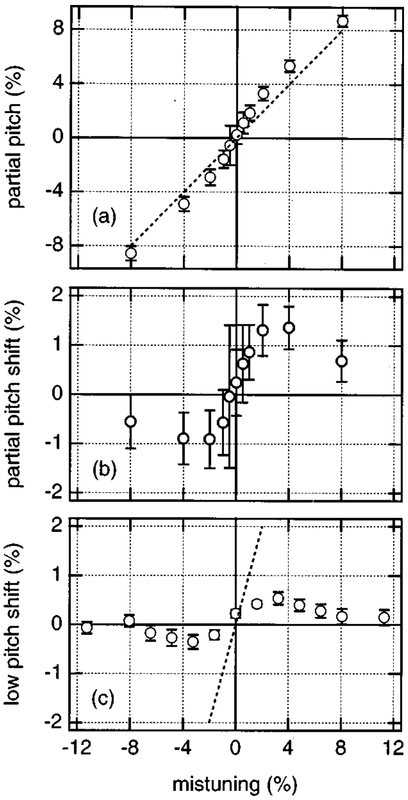

FIG. 1. (a) Pitch of the fifth partial of a complex tone of fundamental 200 Hz as a function of its mistuning. Pitch and mistuning are both expressed as a percentage relative to 1000 Hz. The difference between these data points and the dotted line represents the ''pitch shift of the mistuned partial.'' This shift is replotted in (b). Error bars represent one standard error of the mean (Hartmann and Doty, 1996). (c) Shift of the low pitch of a complex of fundamental 155 Hz as a function of mistuning of the fourth partial (Darwin *et al.*, 1994). The dotted line represents a shift/mistuning slope of one.

However, this description does not match the shifts observed by Hartmann and Doty (1996). The shifts appeared to result not from interactions between neighboring partials, but from the interaction between the partial's frequency, and a *harmonic template* defined by the in-tune partials of the complex (Lin and Hartmann, 1997). A harmonic template was used by the model of Goldstein (1973), to allow noisy estimates of the frequencies of all partials to collectively determine the low pitch of the complex. One might speculate that, in return, such a template somehow affects the pitches of the partials when they are mistuned.

A recent paper proposed a model of such interaction (de Cheveigné, 1997b). A noisy estimate of the partial's frequency determined its pitch, and also the probability of its fusion within the background harmonic complex. This probability varied according to the proximity of the frequency estimate with a harmonic template. If fusion prevents an independent pitch from being heard, it follows that the distribution of pitch matches is biased relative to its distribution in the absence of the complex, and this results in a pitch shift. The model accounted for experimental data quite well, but gave no hint as to the nature of the internal frequency representation, the source of noise, or the nature of the events described by the probability functions. The purpose of the present study is to work out the details of a more concrete model, based on the hypothesis that frequency is represented by the *time-domain* pattern of neural discharge. This hypothesis was supported by Hartmann *et al.* (1990) who found an absolute frequency limit between 2.2 and 3.5 kHz for the segregation of mistuned harmonics, which they interpreted as reflecting a reduction in neural synchrony at high frequencies.

The model described in this paper uses the concept of *harmonic cancellation*, the process by which harmonic (periodic) interference is suppressed to allow targets to be perceived. Harmonic cancellation has emerged experimentally as a principle behind F_0-guided segregation of harmonic sounds (Lea, 1992; Summerfield and Culling, 1992; de Cheveigné *et al.*, 1995, 1997a, 1997b). According to that principle, the auditory system takes advantage of the harmonic structure of interference to get rid of it. Several models of harmonic cancellation have been proposed, among which the channel selection model of Meddis and Hewitt (1992), and the time-domain ''neural cancellation filter'' of de Cheveigné (1993, 1997a). In the latter model, correlates of the interference were suppressed within neural spike trains to allow evidence of the target to emerge. These models explain the improvement in identification observed in ''double-vowel experiments'' when vowels have different fundamental frequencies (F_0) (Scheffers, 1983; Assmann and Summerfield, 1990).

Applied to pitch perception, the ''neural cancellation filter'' allows the period of a stimulus to be estimated from the position of the *minimum* output of an array of such filters, evidence of the best possible cancellation. In this role it is similar to the autocorrelation model of pitch perception (Licklider, 1951; Meddis and Hewitt, 1991a, 1991b), excitatory interaction being replaced by inhibitory interaction, and the search for a maximum by a search for a minimum (de Cheveigné, 1998). The cancellation filter can also assist the estimation of multiple pitches: Tuned to the period of one periodic sound, it suppresses that sound and allows the period(s) of the other sound(s) to be estimated (de Cheveigné, 1993; de Cheveigné and Kawahara, 1999). The present paper is based on the hypothesis that harmonic cancellation is used both in a preprocessing stage, to filter out unwanted components, and in a pitch perception stage to estimate the period. This is shown to result in pitch shifts similar to those observed experimentally.

I. ESTIMATING THE PITCH OF A MISTUNED PARTIAL

The model aims at estimating the pitch of a mistuned partial. It is based on the cancellation model of pitch perception of de Cheveigné (1998) with which it differs in two ways. First, processing is restricted to a single channel (or group of similar channels) tuned to the frequency region of the mistuned partial. Second, an additional branch of neural processing is introduced to weaken correlates of the background harmonic complex and make the mistuned partial easier to estimate. These assumptions are motivated in Sec. I D. Standard and modified cancellation filters are first presented, followed by details of the simulation (spike generation, etc.). They are then applied to the task of pitch perception of a mistuned partial. Finally, the various parameters and assumptions of the model are discussed one after the other.

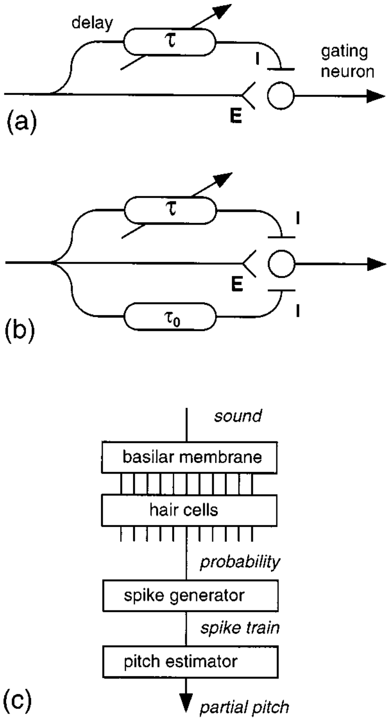

FIG. 2. (a) Standard neural cancellation filter used in the pitch perception model of de Cheveigné (1998). The pitch period is estimated by an array of such filters with delays τ covering the range of expected pitch periods. The delay that produces a minimum output is the period estimate. (b) Modified cancellation filter used in the present model for estimation of a mistuned partial. One delay, τ_0, is fixed and set to the period T_0 of the background harmonic complex. The other, τ, is variable and used to estimate the partial's pitch period T. (c) Schema of the partial pitch perception model. Sound is processed by a gammatone basilar-membrane/haircell model to produce a tonotopic array of auditory-nerve fiber discharge probability functions. A single channel with CF equal to the nominal frequency of the partial is selected to drive a model of spike generation (inhomogenous Poisson process with refractory effects). The spike train is processed by an array of modified cancellation filters (b) to estimate the mistuned partial's pitch.

A. Standard and modified neural cancellation filters

The standard ''neural cancellation filter'' is illustrated in Fig. 2(a). This hypothetical network consists of a gating neuron fed by two pathways, one direct and excitatory, the other delayed and inhibitory. The neuron fires each time a spike arrives along the direct pathway, *unless* a spike arrives simultaneously along the delayed pathway. The neural cancellation filter has been used to account for source segregation on the basis of fundamental frequency (F_0) as observed in ''double-vowel'' experiments (de Cheveigné, 1993, 1997a), and also pitch perception, in a ''negative'' version of the autocorrelation model (de Cheveigné, 1998). To estimate the pitch period, the delay parameter τ is varied until a minimum output is found (equivalently, an array of filters with various τ parameters is scanned). The corresponding delay gives the pitch period estimate. This description applies to a single channel. For the perception of the *low pitch* of a complex, a similar operation is supposed to occur in parallel in all channels emerging from the cochlea. The results of different channels may be combined as suggested by Moore (1997) or Meddis and Hewitt (1991a, 1991b).

For the task of perceiving the pitch of a *single* partial, it is reasonable to suppose that processing is instead restricted to a limited set of channels. In addition, the present model replaces the standard neural cancellation filter of Fig. 2(a) by the ''modified neural cancellation filter'' of Fig. 2(b). In the modified filter, the gating neuron is fed by an extra inhibitory delayed pathway with delay τ_0. Formally, the filter transmits any incoming spike with probability 1 unless the spike is preceded by a spike after an all-order interval in the range $[\tau_0-\Delta\tau_0/2,\tau_0+\Delta\tau_0/2]$ *or* in the range $[\tau-\Delta\tau/2,\tau+\Delta\tau/2]$, in which case transmission probability is 0. Parameters $\Delta\tau_0$ and $\Delta\tau$ determine the width of the ''tolerance windows'' around τ_0 and τ. Particular spikes are removed, and the result is to reduce the density of all-order intervals with durations similar to τ_0 or τ.

When estimating a mistuned partial, the delay τ_0 is set to the period $T_0=1/F_0$ of the background harmonic complex. The motivation for this extra pathway is given in Sec. I D. For now it is sufficient to think of it as a sort of time-domain ''harmonic sieve'' (Duifhuis *et al.*, 1982) that suppresses features in tune with the harmonic series of the background complex, and allows correlates of the mistuned partial to emerge. The second delay, τ, is variable and used to estimate the mistuned partial's period.

B. Simulation

The structure of the model is illustrated in Fig. 2(c). The stimulus waveform (complex with mistuned partial) is filtered by a gammatone filter and hair cell transduction stage (Patterson *et al.*, 1992; Meddis, 1988; Slaney, 1993) that produces an array of discharge probability functions. A single channel is selected, with a characteristic frequency (CF) equal to the nominal frequency of the partial (frequency in the absence of mistuning). It is used to drive a model of auditory nerve fiber discharge production. The resulting spike trains feed the pitch estimation stage that consists of an array of modified cancellation filters with delays covering the range of expected pitch periods.

Stimuli were 16-component complex tones with sine starting phase. The components (partials) followed a 200-Hz harmonic series, with the exception of one partial that could be mistuned by −8% to +8%. Stimuli were generated at a 100-kHz sampling rate and fed to the basilar-membrane/haircell model. Spikes were produced stochastically by an inhomogeneous Poisson process with refractory effects, and stored as interspike intervals with a resolution limited only by the floating point representation. Subsequent calculations used a 10-μs resolution. Refractory effects were modeled as an absolute dead time of 1 ms. Further details and references may be found in de Cheveigné (1998). The average discharge rate was approximately 200 spikes/s. Spikes were produced in parallel for N fibers with similar characteristics ($N=100$). Before pitch estimation, the N spike trains were merged together into a single spike train. This step allows interval statistics to include intervals between spikes of different trains, thus making better use of the available information (de Cheveigné, 1993, 1998). Because of memory limitations, stimuli were shorter than those of Hartmann and Doty (1996): 100 ms vs 400 ms.^{1} The first 20 ms of each spike train were discarded to reduce the effects of ringing and haircell adaptation.

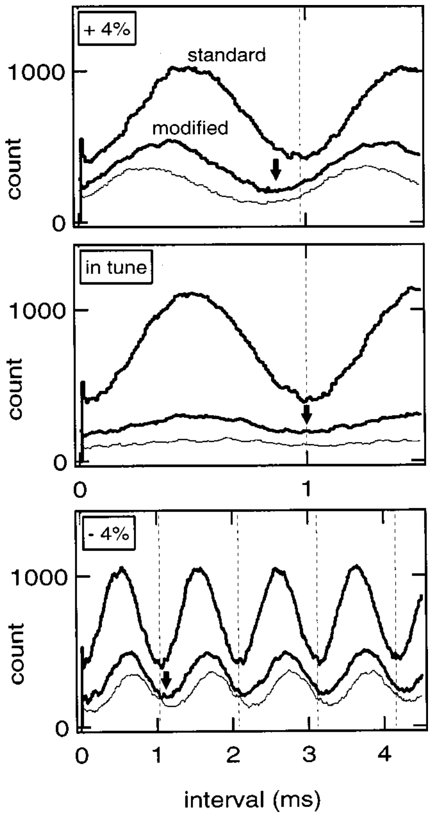

FIG. 3. Top panel, upper thick line: output of a standard cancellation filter as a function of delay, for the fifth partial of a 200-Hz complex. The partial was mistuned by +4%. Lower lines: same for a modified cancellation filter with parameters $\Delta\tau = 50\ \mu s$, and either $\Delta\tau_0 = 50\ \mu s$ (thick line) or $\Delta\tau_0 = 100\ \mu s$ (thin line). Middle panel: same for a mistuning of 0%. Bottom panel: same, for a mistuning of −4%. The horizontal scale was widened to accommodate higher-order minima. Vertical dotted lines indicate integer multiples of the period of the partial.

The model was applied to the mistuned fifth partial of a 200-Hz complex. Figure 3 shows the output of the standard and modified neural cancellation filters for three values of mistuning. In the top panel the partial is mistuned by +4%. The output of the standard cancellation filter (uppermost thick line) shows a minimum as a function of τ at a delay equal to the period (equivalent to the period peak of the autocorrelation model). The function is subject to random variations, but on average the minimum is centered on the period of the partial (dotted line). In contrast, the response of the modified cancellation filter (lower thick line) shows a minimum (arrow) that is shifted in the direction of greater mistunings. That is the origin of the pitch shift in the present model.

In the middle panel the partial is in tune with the background harmonic complex. Neglecting random fluctuations, the period minima of the standard (upper thick line) and modified (lower thick line) filters are centered on the period of the partial (dotted line and arrow). There is no pitch shift. In the bottom panel the partial is mistuned by −4%. The shift in this case is toward longer delays, exaggerating once again the effects of mistuning (arrow). The horizontal scale was widened to show that higher-order minima undergo shifts of the same *absolute* magnitude as the first. Shifts for higher-order minima are thus smaller in proportion, a fact that is exploited later on.

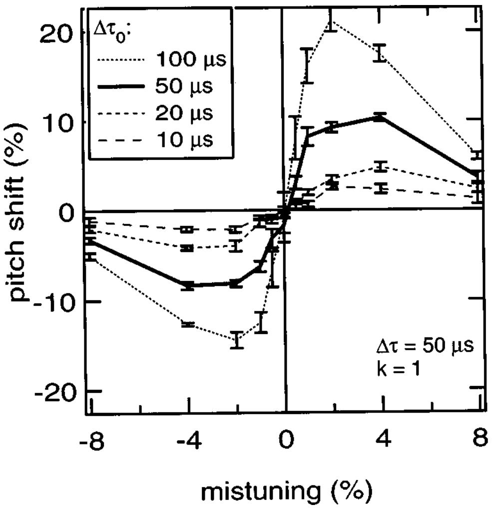

FIG. 4. Pitch shift predicted by the model as a function of mistuning for the fifth partial of a 200-Hz complex, for different values of the $\Delta\tau_0$ parameter. The pitch estimate was based on the position of the first-order minimum ($k = 1$, see text).

Because the extra branch removes more spikes, the output of the modified filter is less than that of the standard filter, particularly when the partial is in tune with the complex (middle panel). The modified filter de-emphasizes in-tune partials. In the simulations that produced the thick lines, $\Delta\tau_0 = \Delta\tau = 0.05$ ms. The thin lines were obtained for $\Delta\tau_0 = 0.1$ ms. With this value, the ''tolerance window'' for removal of intervals equal to T_0 is wider, shifts for the mistuned partials are greater (top and bottom), and the output pattern for the in-tune partial is flatter (middle).

C. Predicted shift patterns

The model was simulated with the same set of mistunings (−8%, −4%, −2%, −1%, −0.5%, 0, 0.5%, 1%, 2%, 4%, 8%) as in the experimental data of Fig. 1(a), (b). For each value of mistuning, the modified cancellation filter was applied to the spike train and τ was varied in a ±40% range around the period of the partial. The array of output spike counts was smoothed by convolution with a 20 point window (0.2 ms) to reduce noise, and the position of the minimum was taken as a period estimate. The simulation was repeated ten times per data point, to refine the averages and estimate their reliability (the purpose was *not* to model the variability of a hypothetical neural process, nor that of subjects responses). Simulations were time and memory consuming, and this limited the range of conditions that could be explored, and the degree to which variability could be reduced by averaging.

Pitch shifts as a function of mistuning for the fifth partial are plotted in Fig. 4. The parameter $\Delta\tau_0$ was given several values, while $\Delta\tau$ was kept fixed at 50 μs. Error bars represent the standard error of the mean over ten simulations. In their *shape*, predicted shift functions are quite similar to the experimental data: Shifts are of the same sign as the mistuning, and they peak at about 4% and decrease beyond. However, predicted shifts are considerably larger than those observed experimentally [Fig. 1(b)]. This problem is addressed in the next section. Shift size varies with $\Delta\tau_0$, which controls the size of the coincidence window for the $\tau_0 = T_0$ branch, and thus indirectly the probability that spikes will be removed by that branch. At $\Delta\tau_0 = 0$ the circuit is equivalent to the standard cancellation filter for which no shifts occur.

D. Questions

So far the model accounts for the shape of the pitch shift function for the fifth partial, but not its magnitude. This section examines that issue and others.

1. Effect size

There are several ways to reconcile the magnitudes of predicted and observed pitch shifts. One is to postulate small values of $\Delta\tau_0$, or a reduced probability of spike suppression in the τ_0 branch of the filter. Another is to suppose that pitch is derived in part from *higher-order* valleys of the pattern (it was seen in Sec. I B that shifts are proportionally smaller for higher-order valleys).

The assumption that higher-order valleys are involved is attractive. The matched filter of Srulovicz and Goldstein (1983) was sensitive to higher-order modes of ISI (first-order interspike interval) histograms. The ''narrowed autocorrelation function'' of Brown and Puckett (1989), which also incorporated information from higher-order modes of the autocorrelation function, was applied to pitch perception by de Cheveigné (1989) and Slaney (1990) to account for the small size of frequency discrimination thresholds and their nearly inverse dependency on frequency and duration (up to 1 kHz and 100 ms, respectively) (Moore, 1973). If this assumption is accepted, simulated shifts can be brought in line with experimental shifts by dividing them by a constant k representing the shift reduction due to the smaller relative shifts of higher-order valleys. With $k = 8$ the match between predicted and experimental data for the fifth partial is good (topmost panel in Fig. 5).

2. Shifts for other partials

Figure 5 shows predicted shifts (lines) for partials 2 to 5, together with experimental data measured by Hartmann and Doty (1996) (symbols). Parameter k was 8, 8, 11, and 8, respectively, for partials 2 to 5. Other model parameters were the same for all partials. Simulated shifts match experimental shifts very well for the fourth and fifth partial, reproducing even the tendency for positive mistunings to produce slightly larger shifts. The match is less good for lower partials, but simulated shifts remain everywhere within experimental error bars. Figure 6 shows similar data for the seventh partial and the fundamental, with k set, respectively, to 8 and 4.

For the seventh partial the experimental data are well predicted for mistunings between −4% and 4%, but not ±8%. At −8% the model predicts a *positive* shift. This is logical because the partial (1288 Hz) is actually closer to the sixth harmonic (1200) than to the seventh (1400). The ''repulsive'' effect of the harmonic series thus increases the pitch estimate instead of decreasing it. A similar explanation holds for the negative shift at 8%. Most other data points of Hartmann and Doty (1996) and Lin and Hartmann (1997) are consistent with the principle of a repulsive effect of the harmonic series, embodied by the model. Experimental data for ±8% mistunings of the seventh partial are exceptions to this principle.

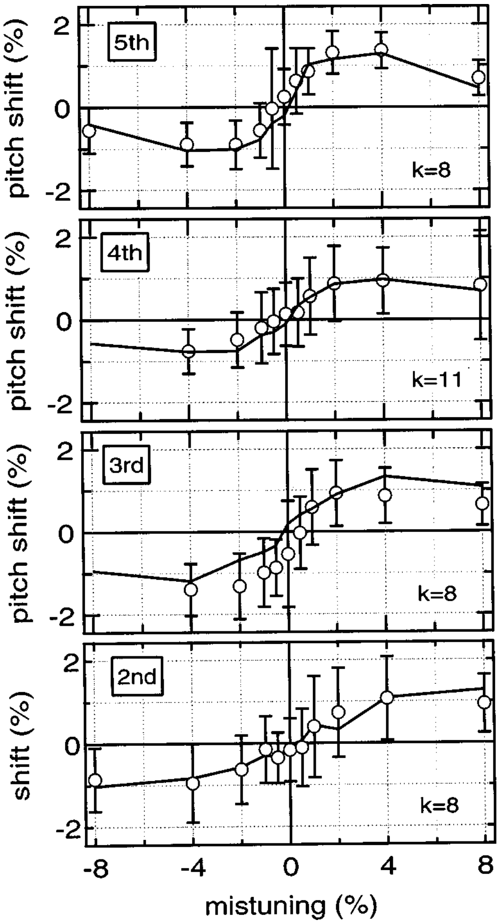

FIG. 5. Lines: pitch shifts predicted by the model as a function of mistuning for partials of various ranks. Symbols: pitch shifts measured experimentally by Hartmann and Doty (1996) for the same partials. Error bars represent one standard error of the mean over experimental data. Parameters $\Delta\tau$ and $\Delta\tau_0$ were the same for all partials. The scaling parameter k was adjusted ''by eye'' to obtain a satisfactory fit.

For the fundamental, the simulated shifts fall within error bars despite a sort of ''glitch'' at 1% and 2% mistuning, that may reflect phase effects as discussed later in this section. Apart from this glitch, the monotonic shape of the experimental shift function at the fundamental is reproduced.

3. Why the extra branch?

Why would the auditory system go to the expense of an extra branch in the cancellation filter, particularly if it produces systematic shifts that the standard filter could avoid? One can suggest that it is part of a mechanism to keep the focus on the mistuned partial, without which attention might wander to neighboring in-tune partials, making the task difficult. The modified filter produces outputs that are larger for mistuned than for in-tune partials, and also more strongly modulated as a function of τ (Fig. 3). Channels responding to the mistuned partial are more prominent, and this eases the task of hearing out the partial. Figure 7 illustrates this point by showing the number of spikes that escape removal by the τ_0 branch as a function of CF, for a mistuning of 4%. Chan-

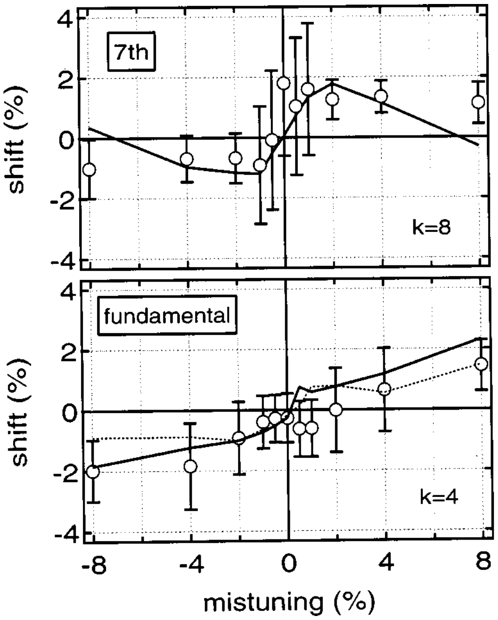

FIG. 6. Same as Fig. 5, for the fundamental and seventh partial. The dotted line in the lower plot (fundamental) is for a stimulus with reversed polarity (180 degree phase shift of all components).

nels responding to the mistuned partial are favored, the contrast being greater for larger values of $\Delta\tau_0$. Such focusing is impossible when the partial is in-tune, which may explain why in-tune partials are harder to hear out.

4. Why not use separate cancellation filters?

The idea of using harmonic cancellation to suppress a competing sound (in this case the background harmonic complex) was put forward in previous studies (de Cheveigné, 1993, 1997a; de Cheveigné and Kawahara, 1999). They assumed, however, that it would take the form of a separate cancellation filter preceding the estimation stage, as illustrated in Fig. 8. Why then postulate the topology of Fig. 2(b)?

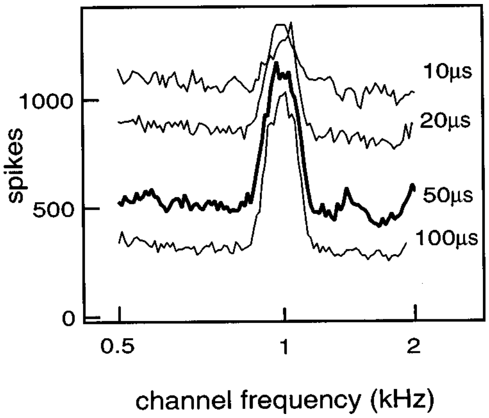

FIG. 7. Number of spikes that escape removal by the τ_0 branch as a function of channel frequency, for various values of the $\Delta\tau_0$ parameter, in response to the 200-Hz complex with a fifth partial mistuned by 4%. More spikes are removed from channels that respond to components in tune with the harmonic complex, than from channels that respond to the mistuned partial. The contrast is greater for larger values of $\Delta\tau_0$.

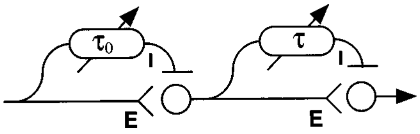

FIG. 8. Two cascaded cancellation filters, as proposed by de Cheveigné and Kawahara (1999) to estimate multiple pitches. The first attenuates components in-tune with one part of the stimulus (for example the harmonic complex), the second is used to estimate the period of the remainder (for example a mistuned partial). This topology is ruled out by the experimental data of Hartmann and Doty (1996), according to the present theory.

It turns out that a model based on the topology of Fig. 8 produces no shifts. To understand why, it is necessary to first understand how shifts were produced by the previous topology. Within a channel tuned to a mistuned partial, spikes tend to occur at peaks of basilar-membrane motion. All-order interspike intervals (ISI) are thus distributed with a maximum corresponding to the period T of the partial. Because of the stochastic nature of spike generation, some spikes occur early within the peak: These tend to be preceded by ISIs that are shorter than T. Likewise, spikes that occur late in the peak tend to be preceded by ISIs that are longer than T. However, both populations are balanced, and the minimum output of a standard cancellation filter is obtained for $\tau = T$.

Consider now the modified cancellation filter [Fig. 2(b)], and suppose that the partial of order n is mistuned positively. In the basilar-membrane region that responds to the partial, peaks of motion occur at intervals T that are shorter than an integer fraction of the period T_0 of the complex. If the τ_0 branch is tuned to T_0, it removes spikes that tend to occur *late* within a peak of basilar-membrane motion. Late-occurring spikes being removed by the τ_0 branch, it is advantageous to mistune the τ branch to remove spikes that occur *early*, because that maximizes spike suppression. The minimum as a function of τ is thus displaced. That is the origin of the shift with the filter of Fig. 2(b).

In the case of cascaded filters as in Fig. 8, the second cancellation filter (tuned to τ) estimates the most common interval in the output of the first. Interval statistics are modified by the first filter, that removes certain spikes, but this does not result in the systematic shifts observed with the other topology. To understand why, notice that every spike determines two populations of intervals—those that *end* on the spike, and those that *begin* on it. If a spike occurs late, one population is biased toward long intervals, but the other is equally biased toward short intervals. The net effect of removing that spike on average interval length is zero. The interval statistics cannot be significantly biased in this way, and that explains the lack of shifts for the topology of Fig. 8.

The lack of shifts for the circuit of Fig. 8 explains why that topology was not chosen for the pitch shift model. It does not explain why the auditory system might have made the same choice. Figure 8 requires an extra relay with high temporal accuracy. Not only that, but each relay requires its own array of delay lines, whereas in Fig. 2(b) both delay lines are chosen from the same array. Considerations of this type may explain why the auditory system adopted the topology of Fig. 2(b), despite the fact that it is conceptually more complex and produces pitch shifts. In any case, it is remarkable that the psychophysical data of Hartmann and Doty (1996) can be translated into such a precise anatomical constraint.

5. Why cancellation?

Cancellation and autocorrelation are closely related (de Cheveigné, 1998). Autocorrelation being familiar to many, it is reasonable to ask whether the model can be recast in terms of a ''modified autocorrelation,'' by replacing inhibitory by excitatory interaction in the circuit of Fig. 2(b). The answer is yes, in the sense that reversing the sign of the interaction produces a ''modified autocorrelation function'' with maxima that show exactly the same shifts. However, whereas the extra inhibitory branch made sense in terms of cancellation of an unwanted background complex, it is harder to justify the extra branch if it is excitatory. Across channels, the output would be *smaller* in channels tuned to the mistuned partial instead of larger, which makes little sense if the task is to hear out that partial. Cancellation is more logical, but one should not reject the possibility that the shifts occur within a different kind of neural circuit, for example excitatory.

6. How is T_0 estimated?

The T_0 estimate (necessary to set the fixed delay τ_0) may be derived from the period estimation stage of any model of low pitch perception, for example, autocorrelation or cancellation. Judging from the shifts of the low pitch as a function of mistuning [Fig. 1(c)], it would appear that channels responding to the mistuned partial are included in the estimate for small mistunings (<4%) and excluded for large mistunings (>4%). At small mistunings the T_0 estimate might thus be slightly shifted, but this is of little consequence to the model as long as the shift is small relative to the mistuning.

What range of channels must be included in the T_0 calculation? Lin and Hartmann (1997) found shifts for harmonic complexes that lacked many partials, in particular those immediately above and/or below the mistuned partial. This suggests that the estimation of T_0 involves a broad range of channels. On the other hand, Roberts and Brunstrom (1998) found shifts for partials mistuned from a slightly inharmonic complex (derived from a harmonic complex by uniformly ''shifting'' or ''stretching'' it). This suggests instead that the template (or T_0 estimate) is derived from a set of partials neighboring the mistuned partial. More research is needed to resolve this apparent contradiction. The idea of a ''local'' harmonic template parallels the ''modified Equalization Cancellation'' model of Culling and Summerfield (1995; Culling *et al.*, 1998a, 1998b) in which binaural cancellation occurs with parameters local to each channel, rather than common to all channels as in the original EC model of Durlach (1963). Local harmonic templates might also explain monaural segregation data for vowels with formants or partials exchanged between harmonic series in a way that would confuse a mechanism that fits all harmonics of a vowel to a single harmonic series (Culling and Darwin, 1993, 1994).

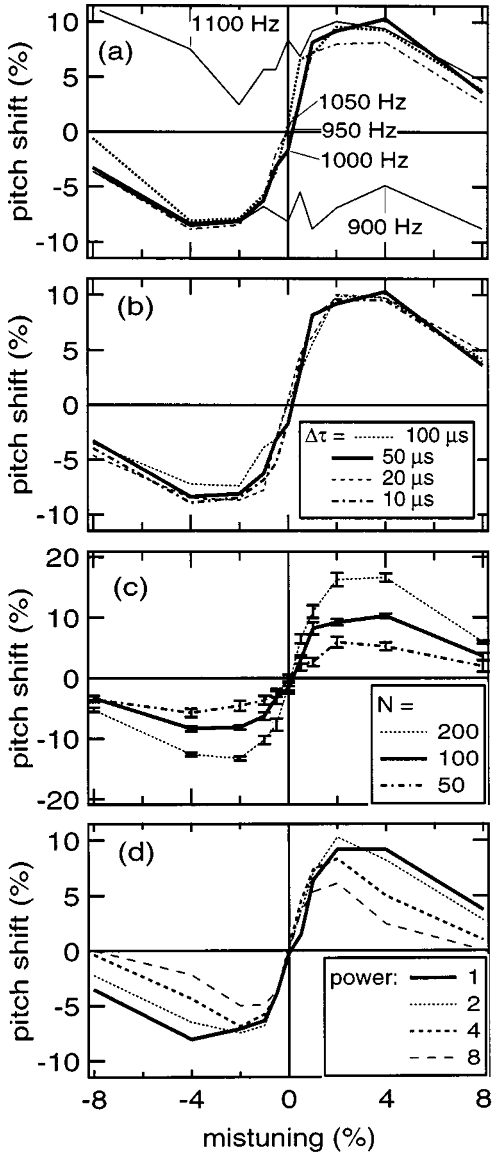

FIG. 9. Effects of various model parameters on the pitch shift as a function of mistuning of the fifth partial of a 200-Hz complex. (a) Effect of the channel frequency. Each line is for a different channel of the basilar-membrane/haircell model ($k=1$, $\Delta\tau=\Delta\tau_0=50\ \mu s$, $N=100$). (b) Effect of $\Delta\tau$ ($k=1$, $\Delta\tau_0=50\ \mu s$, $N=100$). (c) Effect of the number of fibers N ($k=1$, $\Delta\tau=\Delta\tau_0=50\ \mu s$). (d) Effect of adding a sharpening stage, implemented by raising the haircell model output to a power before spike generation. Each line is for a different power (power$=1$ implies no sharpening) ($k=1$, $\Delta\tau=\Delta\tau_0=50\ \mu s$, $N=100$).

7. Effects of other parameters

Figure 9(a) shows pitch shifts obtained from five different channels of the basilar-membrane/haircell model (900 Hz, 950 Hz, 1000 Hz, 1050 Hz, 1100 Hz), as a function of mistuning of the fifth partial. Estimates are erratic for the two extreme channels, but consistent between the three central channels, and presumably any other channels within this range. Within this range, the estimates do not depend on which set of channels are included in the simulation. Based on the cochlear frequency map of Greenwood (1990), and an estimated number of 30 000 auditory nerve fibers in man, the cochlear region between 950 and 1050 Hz feeds about 540 fibers, an ample number for the model. With a larger value of $\Delta\tau_0$ (more severe rejection of the background harmonic series), estimates are consistent over an even wider range of channels (not shown).

Figure 9(b) shows the effect of the parameter $\Delta\tau$. $\Delta\tau_0$ was kept constant at 50 μs. The effect of this parameter is small.

Figure 9(c) shows the effect of varying the number *N* of fibers used in the simulation. Shift size is reduced when *N* is reduced, but the shape of the shift function hardly changes. A change in discharge rate (for example, due to a change in stimulus level) would have a similar effect.

It has been suggested that pitch perception might involve *temporal sharpening* of neural discharge patterns, for example, by onset cells in the cochlear nucleus. Sharpening was simulated by raising to a power the probability function at the output of the haircell model, and scaling it to obtain an average probability of 200 spikes/s. Effects of sharpening are displayed in Fig. 9(d). Sharpening causes the shift pattern to vary over a narrower range, while the slope near 0% mistuning remains approximately the same. A close match to experimental data was found without sharpening (Fig. 5). It can only be degraded if sharpening is introduced. This argues *against* the hypothesis that auditory nerve fiber discharge patterns undergo temporal sharpening before pitch extraction, at least for this particular task (see also Shofner, 1998).

In summary, the model is remarkably insensitive to changes in parameters, other than the simple scaling effect of changing $\Delta\tau_0$, *k*, or *N*. To change the shape of the shift function required a structural change (addition of a sharpening stage) that the experimental data do not support. The fact that the model fits the data *despite* this lack of flexibility can be interpreted as an additional argument in favor of the model.

8. Evidence of phase effects

Mistuning causes the stimulus waveform to beat at a rate of nmF_0, where *n* is the rank of the mistuned partial, *m* is the mistuning factor, and F_0 is the fundamental. For the fifth partial of a 200-Hz complex, the beat periods for mistunings of ±0.5%, ±1%, ±2%, ±4%, ±8% are, respectively, 200 ms, 100 ms, 50 ms, 25 ms, and 12.5 ms. Beat periods are proportionally longer for mistuned partials of lower ranks (for the fundamental they are, respectively, 1000 ms, 500 ms, 250 ms, 125 ms, and 62.5 ms). The waveform fluctuates with the beat, and the shape of a stimulus thus depends on which part of the beat period it samples, which in turn depends on the relative starting phases of the various components. Phase effects are thus to be expected. They should be small when the beat period is shorter than the stimulus, but possibly relatively large when the beat period is longer. Indeed, Hartmann (1988) found phase effects for the detection of mistuned partials, particularly for short stimuli.

The present simulations used 100-ms stimuli (the first 20 ms of the spike trains produced were removed, so the model operated essentially on the last 80 ms). This is shorter than a beat period for many conditions, particularly for partials of low rank. The minor irregularities visible in Fig. 5 (second harmonic) and Fig. 6 (fundamental) are highly reproducible, and it turns out that they are also phase dependent. The dotted line in Fig. 6 (fundamental) was obtained for a reversal in polarity of the stimulus (180 degree phase shift of all components). There is thus some evidence that the model is phase sensitive. However, no attempt was made to extend it to account for the phase dependency of detection reported by Hartmann (1988).

II. DISCUSSION

A previous paper (de Cheveigné, 1997b) proposed a different model to explain the same phenomena. That model did not suggest what physiological mechanisms were involved. In particular it was neutral on the question of whether pitch is derived from time-domain or spectral (place) representations. The present paper offers a more concrete model based on the hypothesis of time-domain processing of neural spike trains. In the absence of an equally plausible place model, this theory (together with the phenomenon that it explains) can be interpreted as supporting a time-domain model of pitch perception.

The model is congruent with the hypothesis of a time-domain neural cancellation filter that was proposed to explain concurrent voice segregation (de Cheveigné, 1993, 1997a) and pitch perception (de Cheveigné, 1998), and in particular with the hypothesis that harmonic cancellation is used to ease the perception of pitch in the presence of competing harmonic sounds (de Cheveigné and Kawahara, 1999). However, the lack of shifts for the topology of Fig. 8 suggests that the arbitrary levels of cascaded filters imagined by de Cheveigné and Kawahara (1999) might not be available in the auditory system. The theory puts strong constraints on the topology of the neural circuits involved in this task, by ruling out one topology that was *a priori* plausible, and showing that another can account for the shifts. However, it does *not* prove that it is the only one. In particular it was pointed out that replacing inhibitory by excitatory interaction in Fig. 2(b) would result in similar shifts.

The question arises as to what part of the auditory system might house such processing. The auditory system contains circuitry specialized in relaying high-resolution temporal information. Globular and spherical bushy cells in the cochlear nucleus (CN) are fed by axons of the auditory nerve via secure synapses that guarantee little degradation of temporal accuracy. They project to a number of relays in the brainstem: contralateral CN, superior olivary complex, nuclei of the lateral lemniscus (NLL), and inferior colliculus (IC) (Rhode and Greenberg, 1992; Romand and Avan, 1997). Within the superior olivary complex, principal cells in the medial nucleus of the trapezoidal body (MNTB) are fed by thick axons from contralateral globular bushy cells, via specialized synapses called calyces of Held. These neurons, which are inhibitory, project to ipsilateral lateral superior olive (LSO) and are supposed to be involved in the processing of interaural level differences (ILD). ILDs in principle do not involve time, and it is therefore surprising to find time-specialized circuitry involved in their processing. It has been suggested that the LSO also processes onset time disparities

(Batra *et al.*, 1997; Joris and Yin, 1998). One can speculate that this circuitry supports other forms of time-domain processing. In addition to the ipsilateral LSO, principal cells of the MNTB project to the ventral nucleus of the lateral lemniscus (VNLL) and to dorsomedial and ventromedial periolivary nuclei (DMPO and VMPO) (Spangler *et al.*, 1985). Smith *et al.* (1998) cite as an enigma that these regions receive both excitatory and inhibitory input from contralateral bushy cells, an arrangement that makes little sense in terms of binaural processing.

Other projections of the MNTB include the IC, contralateral CN, and even the outer haircells of the (mainly contralateral) cochlea, via the medial olivo-cochlear bundle (MOCB) (Helfert and Aschoff, 1997; Irvine, 1992; Warr, 1992). The hypothesis that cancellation occurs in the mechanical domain, involving reverse transduction of time-domain signals carried by efferent pathways, is attractive given the relative linearity and wide dynamic range of mechanical vibration. Unfortunately that hypothesis seems unlikely, as recordings from the MOCB have failed to reveal any stimulus-locked temporal structure (Guinan, 1996). In brain-slice preparations of the mouse CN, Wickesberg and Oertel (1990) found evidence for delayed inhibitory input to bushy and stellate cells in the anteroventral cochlear nucleus (AVCN) via the dorsal cochlear nucleus (DCN). They suggested that it served for monaural echo suppression, but harmonic cancellation might be another candidate role.

Neurons with the right sort of gating properties have been found in the LSO (lateral superior olive) (Batra *et al.*, 1997; Joris *et al.*, 1998; Sanes, 1990). In those neurons, presynaptic discharges in excitatory afferents produce EPSPs (excitatory post-synaptic potentials) and those in inhibitory afferents produce IPSPs (inhibitory post-synaptic potentials), both of which appear to be of short duration in certain LSO cells (Batra *et al.*, 1997). Supposing that an EPSP is sufficient to reach the neuron's firing threshold, and a well-timed IPSP sufficient to *prevent* the threshold from being reached, the neuron's properties would match those postulated by the neural cancellation filter.

The lack of evidence for delay lines has been pointed out many times. It is, however, possible that such delays might be hard to observe, if they occur within thin axonal fibers that feed the processing neuron. An *array* of neurons ordered according to delay or period might not be observed, if the delays were tuned, for example, by switching afferents.

Maas (1997a, 1997b) gave two theorems that support the thesis that spike-time representations may be used for complex processing in the nervous system (auditory or otherwise). The first demonstrates that any network based on a rate representation processed by ''sigmoidal'' neurons can be simulated by a network of spiking neurons, of comparable size. The second shows that for certain tasks (for example, the ''winner take all'' task) a sigmoidal network requires many more neurons than a spiking network. Together they imply that spiking neurons are more ''powerful'' than sigmoidal neurons. Spike-time processing is also much faster. These considerations argue against the proposition that a temporal code, which is natural in the auditory system, must necessarily be translated to a rate/place code before significant processing can occur. Maas has also suggested that filters of arbitrary impulse response could be ''synthesized'' by combinations of EPSPs and IPSPs, each weighted by the strength of the synapse and time-shifted by the delay of the pathway that feeds it. The circuits of Fig. 2 are simple examples of that idea, but one could speculate on more sophisticated circuits for the processing of vowel timbre, or dynamic features of speech or environmental sounds.

The model is specialized for a particular task. The assumptions it makes (an extra delayed branch, selection of a subset of channels) would make no sense if the task were, say, to match the low pitch of a harmonic complex. The two tasks therefore require different models. One can object that this violates the principle of parsimony, yet it is not unreasonable to imagine that the auditory system might employ different tools to solve different problems. To segregate a high-order partial from a harmonic complex, and match its pitch, is a more complex operation than to simply match the low pitch. There is a striking gap between the complexity of anatomic and physiological descriptions of the auditory system, and the models available to account for them. By maintaining that gap, an overdogmatic defense of parsimony might do a disservice to our understanding. It helps that various models use a uniform set of ingredients: delay, coincidence (whether excitatory or inhibitory), temporal integration, selective weighting (for example, of different tonotopic channels), etc. It would help further still if they could be assembled in a modular fashion, for example, by placing the pitch estimation model at the output of the module that suppresses the background complex. By ruling out the topology of Fig. 8 in favor of that of Fig. 2(b), the present study showed the limits of such modularity.

III. CONCLUSIONS

(1) The systematic shifts in the pitch of a partial that are observed when that partial is mistuned from a harmonic complex can be reproduced within a model based on time-domain processing of auditory nerve fiber discharge patterns.

(2) The model involves a neural circuit with a gating neuron fed by three pathways, one direct and excitatory, and the other two delayed and inhibitory. The delay parameter of one of the delayed pathways is varied in search of a minimum. The delay at the minimum is an estimate of the partial's period, and is used as the cue to the pitch of the partial in the model.

(3) The delay of the second pathway is tuned to the period of the background harmonic complex. Its purpose is to suppress the in-tune correlates of the complex, and thus ease the task of hearing out the mistuned partial.

(4) The pitch shift arises as a result of the stochastic nature of nerve discharge production, and its interaction with the two delayed pathways.

(5) The shifts are absent in a circuit that is *a priori* at least as plausible, consisting of two independent neural cancellation filters. The existence of shifts thus suggests constraints on the topologies of neural processing available for this kind of task.

(6) The fact that a plausible mechanism of time-domain neural processing predicts the same pitch shifts as observed experimentally argues in favor of a time-domain mechanism of pitch perception, at least for the pitch of partials. The weight of this argument depends upon the lack of a similarly detailed model of the spectral (place) variety.

ACKNOWLEDGMENTS

Part of this work was carried out at ATR Human Information Processing Research Laboratories, within a research agreement with the Center National de la Recherche Scientifique and the University of Paris 7. The author thanks ATR for its kind hospitality, and the CNRS for leave of absence. Chris Darwin provided the data for Fig. 1(c) and Bill Hartmann that of Figs. 5 and 6. The simulations relied heavily on Malcolm Slaney's AuditoryToolbox code. Thanks to Bill Hartmann, two anonymous reviewers, and the editor for useful comments on previous versions of the manuscript.

---

[1] The shorter duration (100 ms instead of 400 ms) increases the sensitivity of the simulation to phase effects that can occur when the stimulus is shorter than a beat period. Phase effects were observed, particularly at the fundamental, as discussed in Sec. I D. Based on simulations with various starting phase spectra, it appears that they are not large enough to affect the major pattern of simulated shifts, or the validity of the conclusions.

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

Batra, R., Kuwada, S., and Fitzpatrick, D. C. (**1997**). ''Sensitivity to interaural temporal disparities of low- and high-frequency neurons in the superior olivary complex. II. Coincidence detection,'' J. Neurophysiol. **78**, 1237–1247.

Brown, J. C., and Puckette, M. S. (**1989**). ''Calculation of a ''narrowed'' autocorrelation function,'' J. Acoust. Soc. Am. **85**, 1595–1601.

Culling, J. F., and Darwin, C. J. (**1993**). ''Perceptual separation of simultaneous vowels: Within and across-formant grouping by $F0$,'' J. Acoust. Soc. Am. **93**, 3454–3467.

Culling, J. F., and Darwin, C. J. (**1994**). ''Perceptual and computational separation of simultaneous vowels: Cues arising from low frequency beating,'' J. Acoust. Soc. Am. **95**, 1559–1569.

Culling, J. F., and Summerfield, Q. (**1995**). ''Perceptual segregation of concurrent speech sounds: Absence of across-frequency grouping by common interaural delay,'' J. Acoust. Soc. Am. **98**, 785–797.

Culling, J. F., Marshall, D., and Summerfield, Q. (**1998b**). ''Dichotic pitches as illusions of binaural unmasking II: The Fourcin pitch and the dichotic repetition pitch,'' J. Acoust. Soc. Am. **103**, 3527–3539.

Culling, J. F., Summerfield, Q., and Marshall, D. H. (**1998a**). ''Dichotic pitches as illusions of binaural unmasking I: Huggin's pitch and the 'binaural edge pitch,' '' J. Acoust. Soc. Am. **103**, 3509–3526.

Darwin, C. J., Ciocca, V., and Sandell, G. J. (**1994**). ''Effects of frequency and amplitude modulation on the pitch of a complex tone with a mistuned harmonic,'' J. Acoust. Soc. Am. **95**, 2631–2636.

de Cheveigné, A. (**1989**). ''Pitch and the narrowed autocoincidence histogram,'' Proc. ICMPC, Kyoto, 67–70.

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

de Cheveigné, A., and Kawahara, H. (**1999**). ''Multiple period estimation and pitch perception model,'' Speech Commun. **27**, 175–185.

de Cheveigné, A., Kawahara, H., Tsuzaki, M., and Aikawa, K. (**1997a**). ''Concurrent vowel identification I: Effects of relative level and $F0$ difference,'' J. Acoust. Soc. Am. **101**, 2839–2847.

de Cheveigné, A., McAdams, S., and Marin, C. (**1997b**). ''Concurrent vowel identification II: Effects of phase, harmonicity and task,'' J. Acoust. Soc. Am. **101**, 2848–2856.

de Cheveigne, A., McAdams, S., Laroche, J., and Rosenberg, M. (**1995**). ''Identification of concurrent harmonic and inharmonic vowels: A test of the theory of harmonic cancellation and enhancement,'' J. Acoust. Soc. Am. **97**, 3736–3748.

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

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

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., Gerson, A., Srulovicz, P., and Furst, M. (**1978**). ''Verification of the optimal probabilistic basis of aural processing in pitch of complex tones,'' J. Acoust. Soc. Am. **63**, 486–497.

Greenwood, D. D. (**1990**). ''A cochlear frequency-position function for several species—29 years later,'' J. Acoust. Soc. Am. **87**, 2592–2605.

Guinan, J. J. (**1996**). ''Physiology of olivocochlear efferents,'' in *The Cochlea*, edited by P. Dallos, A. N. Popper, and R. R. Fay (Springer Verlag, New York), pp. 435–502.

Hartmann, W. M., McAdams, S., and Smith, B. K. (**1986**). ''Matching the pitch of a mistuned harmonic in a complex tone,'' IRCAM technical report.

Hartmann, W. M. (**1988**). ''Pitch perception and the segregation and integration of auditory entities,'' in *Auditory Function—Neurological Bases of Hearing*, edited by G. M. Edelman, W. E. Gall, and W. M. Cowan (Wiley, New York), pp. 623–645.

Hartmann, W. M., McAdams, S., and Smith, B. K. (**1990**). ''Hearing a mistuned harmonic in an otherwise periodic complex tone,'' J. Acoust. Soc. Am. **88**, 1712–1724.

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.

Helfert, R. H., and Aschoff, A. (**1997**). ''Superior olivary complex and nuclei of the lateral lemniscus,'' in *The Central Auditory System*, edited by G. Ehret and R. Romand (Oxford University Press, New York), pp. 193–258.

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

Hukin, R. W., and Darwin, C. J. (**1995**). ''Comparison of the effect of onset asynchrony on auditory grouping in pitch matching and vowel identification,'' Percept. Psychophys. **57**, 191–196.

Irvine, D. R. F. (**1992**). ''Physiology of the auditory brainstem,'' in *The Mammalian Auditory Pathway: Neurophysiology*, edited by A. N. Popper and R. R. Fay (Springer Verlag, New York), pp. 153–231.

Joris, P. X., and Yin, T. C. T. (**1998**). ''Envelope coding in the lateral superior olive. III. Comparison with afferent pathways,'' J. Neurophysiol. **79**, 253–269.

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.

Lin, J. L., and Hartmann, W. M. (**1997**). ''The pitch of mistuned harmonics: Evidence for a template model,'' J. Acoust. Soc. Am. **103**, 2608–2617.

Maas, W. (**1997a**). ''Networks of spiking neurons: the third generation of neural network models,'' Neural Networks **10**, 1659–1671.

Maas, W. (**1997b**). ''Fast sigmoidal networks via spiking neurons,'' Neural Comput. **9**, 279–304.

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: 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 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. (**1973**). ''Frequency difference limens for short-duration tones,'' J. Acoust. Soc. Am. **54**, 610–619.

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

Moore, B. C. J. (**1987**). ''The perception of inharmonic complex tones,'' in *Auditory Processing of Complex Sounds*, edited by Y. A. Watson (Laurence Erlbaum, Hillsdale), pp. 180–189.

Moore, B. C. J., Peters, R. W., and Glasberg, B. R. (**1985**). ''Thresholds for the detection of inharmonicity in complex tones,'' J. Acoust. Soc. Am. **77**, 1861–1867.

Moore, B. C. J., Peters, R. W., and Glasberg, B. R. (**1986**). ''Thresholds for hearing mistuned partials as separate tones in harmonic complexes,'' J. Acoust. Soc. Am. **80**, 479–483.

Patterson, R. D., Robinson, K., Holdsworth, J., McKeown, D., Zhang, C., and Allerhand, M. (**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.

Rhode, W. S., and Greenberg, S. (**1992**). ''Physiology of the cochlear nuclei,'' in *The Mammalian Auditory Pathway: Physiology*, edited by A. N. Popper and R. R. Fay (Springer Verlag, New York), pp. 94–152.

Roberts, B., and Brunstrom, J. M. (**1998**). ''Perceptual segregation and pitch shifts of mistuned components in harmonic complexes and in regular inharmonic complexes,'' J. Acoust. Soc. Am. **104**, 2326–2338.

Romand, R., and Avan, P. (**1997**). ''Anatomical and functional aspects of the cochlear nucleus,'' in *The Central Auditory System*, edited by G. Ehret and R. Romand (Oxford University Press, New York), pp. 97–191.

Sanes, D. H. (**1990**). ''An in vitro analysis of sound localization mechanisms in gerbil lateral superior olive,'' J. Neurosci. **10**, 3494–3506.

Scheffers, M. T. M. (**1983**). ''Sifting vowels,'' Gröningen unpublished doctoral dissertation.

Shofner, W. P. (**1998**). ''Evidence that primarylike units are the major CN subsystem which encodes pitch related information in their temporal discharge,'' ARO abstract #382.

Slaney, M. (**1990**). ''A perceptual pitch detector,'' Proc. ICASSP-90, pp. 357–360.

Slaney, M. (**1993**). ''An efficient implementation of the Patterson-Holdsworth auditory filter bank,'' Apple Computer technical report, 35.

Smith, P. H., Joris, P. X., and Yin, T. C. T. (**1998**). ''Anatomy and physiology of principal cells of the medial nucleus of the trapezoid body (MNTB) of the cat,'' J. Neurophysiol. **79**, 3127–3142.

Spangler, K. M., Warr, B., and Henkel, C. K. (**1985**). ''The projections of principal cells of the medial nucleus of the trapezoid body in the cat,'' J. Comp. Neurol. **238**, 249–262.

Srulovicz, P., and Goldstein, J. L. (**1983**). ''A central spectrum model: A synthesis of auditory-nerve timing and place cues in monaural communication of frequency spectrum,'' J. Acoust. Soc. Am. **73**, 1266–1276.

Summerfield, Q., and Culling, J. F. (**1992**). ''Periodicity of maskers not targets determines ease of perceptual segregation using differences in fundamental frequency,'' Proc. 124th meeting of the ASA, 2317(A).

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

Terhardt, E. (**1979**). ''Calculating virtual pitch,'' Hearing Res. **1**, 155–182.

Warr, W. B. (**1992**). ''Organization of olivocochlear efferent systems in mammals,'' in *The Mammalian Auditory Pathway: Neuroanatomy*, edited by D. B. Webster, A. N. Popper, and R. R. Fay (Springer Verlag, New York), pp. 410–448.

Wickesberg, R. E., and Oertel, D. (**1990**). ''Delayed, frequency-specific inhibition in the cochlear nuclei of mice: A mechanism for monaural echo suppression,'' J. Neurosci. **10**, 1762–1768.

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