Virtual pitch and phase sensitivity of a computer model of the auditory periphery. II: Phase sensitivity

Ray Meddis and Michael J. Hewitt
Department of Human Sciences, University of Technology, Loughborough LE11 3TU, United Kingdom

(Received 18 September 1990; revised 6 November 1990; accepted 24 January 1991)

In a companion article [Meddis and Hewitt, J. Acoust. Soc. Am. **89**, 2866–2882 (1991)] it was shown that a computational model of the auditory periphery followed by a system of autocorrelation analyses was able to account for a wide range of human virtual pitch perception phenomena. In this article it is shown that the same model, with no substantial modification, can predict a number of results concerning human sensitivity to phase relationships among harmonic components of tone complexes. The model is successfully evaluated using (a) amplitude-modulated and quasifrequency-modulated stimuli, (b) harmonic complexes with alternating phase change and monotonic phase change across harmonic components, and (c) mistuned harmonics. The model is contrasted with phase-insensitive theories of low-level auditory processing and offered as further evidence in favor of the value of analysing time intervals among spikes in the auditory nerve when explaining psychophysical phenomena.

PACS numbers: 43.66.Nm, 43.66.Ba, 43.66.Hg [WAY]

INTRODUCTION

In a previous article (Meddis and Hewitt, 1991), we demonstrated that a peripheral auditory model based on an analysis of time intervals among spikes in the auditory nerve could simulate many aspects of human pitch perception. In this article, we aim to show that the same model can mimic a number of important aspects of human listeners' sensitivity to phase. The issue of phase sensitivity is critical to the discussion of theories of low-level auditory perception because the theories are readily divided into two groups: those that do and those that do not admit of any sensitivity to phase.

In this context, "phase" normally refers to the phase relationships among a set of otherwise harmonically related tone components. There is little evidence that listeners are sensitive to any other aspect of stimulus phase. Indeed, it has been widely believed for some time that the human ear is phase insensitive. However, evidence has accumulated in recent years (Ritsma and Engel, 1964; Moore, 1977; Patterson, 1987; Hartmann, 1988; Sivaramakrishnan *et al.*, 1989) that the phase relationships of tone components can produce clearly perceptible effects under certain circumstances. Nevertheless, there are many situations in which the manipulation of phase leads to no noticeable effect. It is an important test of any model that it should be phase sensitive for certain stimuli but phase insensitive for others.

Some place theories of pitch perception (Goldstein, 1973; Wightman, 1973b; Terhardt, 1974, 1979) are characterized by a peripheral spectral analysis of the acoustic signal that discards phase information. Buunen *et al.* (1974) presented a place theory that did take phase into account by postulating that the phase relationships among harmonics determined the strength of the combination tones and hence the timbre of the whole complex. Unfortunately, the theory was not fully developed and tested against a wide range of stimuli. Competing temporal theories (e.g., Schouten, 1970; Licklider, 1951; Bilsen and Ritsma, 1969; Moore, 1982) imply that signal analysis occurs in the temporal domain which preserves aspects of the fine structure of the peripherally bandpass-filtered signal, at least at low and medium stimulus frequencies. Because this fine structure contains signal phase information, the issue of phase sensitivity has been central to the controversy concerning these two opposing perspectives.

The argument has centered on the ability of phase changes to alter the pitch of a stimulus (Ritsma and Engel, 1964; Lundeen and Small, 1984; Moore, 1977; Wightman, 1973a) and we shall examine some of the tests of this possibility below. The general issue of phase sensitivity, however, is much broader and includes timbre changes as demonstrated by Patterson (1987) using harmonic complexes consisting of components with variable phase. In addition, short stimuli consisting of harmonic complexes with a single, slightly ($< 10\%$) mistuned harmonic can be characterized as having a single component with phase gradually advancing with respect to the other components. Hartmann (1988) has shown that a listener's ability to discriminate such a stimulus from a harmonic complex is tightly correlated with the momentary relative phase of the mistuned harmonic. We shall examine the ability of the model to simulate some of these important results.

I. MODEL DESCRIPTION

The model outlined in Fig. 1 consists of eight stages; each of these stages has been described individually and in detail in our previous article (Meddis and Hewitt, 1991). The current implementation was exactly the same except for an additional stage which attempts to predict the discriminability of two stimuli (see stage eight, below). The stages are as follows:

(1) outer-ear low- and high-frequency attenuation.

(2) middle-ear low- and high-frequency attenuation.

 0001-4966/91/062883-12$00.80

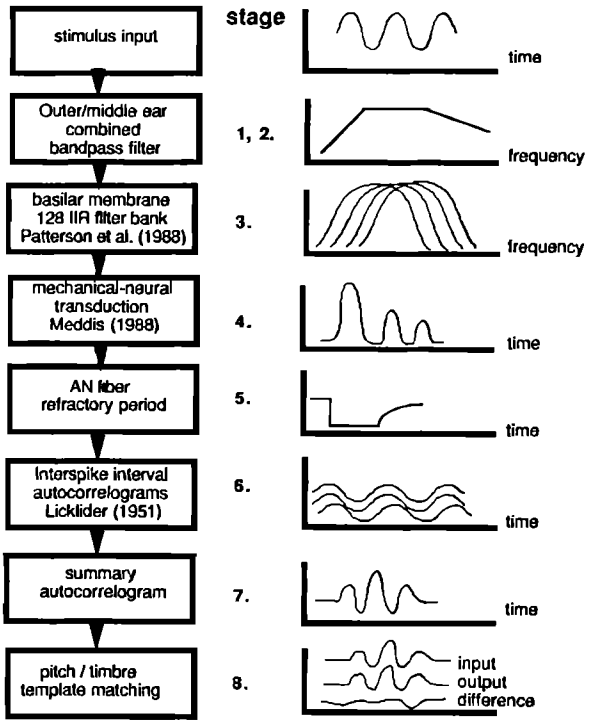

FIG. 1. Processing sequence of the model.

Stages 1 and 2 produce a combined attenuation of frequencies below 1 kHz and above 5 kHz.

(3) inner-ear, cochlear mechanical filtering of the basilar membrane. This was implemented using 128 overlapping bandpass digital filters equally spaced on an ERB-rate scale between 80 Hz and 8 kHz. The filters, which are linear and symmetrical, were supplied to us by John Holdsworth and Roy Patterson at the Applied Psychology Research Unit, Cambridge, UK (Patterson *et al.*, 1988).[1] The theory behind these rounded-exponential (roex) filter shapes is given in Patterson and Moore (1986).

(4) mechanical to neural transduction at the hair cell. This was accomplished using a hair cell model described in Meddis (1986, 1988) and Meddis *et al.* (1990) and the results expressed as the probability of occurrence of a spike in the auditory-nerve fibers of the 128 channels.

(5) refractory inhibition of firing of auditory-nerve fibers. This was computed as an adjustment to the fiber firing probability as a function of the time since it last generated a spike.

(6) short-term estimation of the distribution of intervals among all spikes coming from fibers within the same channel. This is equivalent to the evaluation of a running autocorrelation function (ACF) with a short time constant (in our case 2.5 ms):

$$h(t,\delta t,k) = \sum_{i=1}^{\infty} p(t-T)\, p(t-T-\delta t)\, e^{-T/\Omega}\, dt, \quad (1)$$

where k is the channel number, dt is the sample period, δt is the autocorrelation lag, $p(t)$ is the probability of a spike between time t and $t + dt$, Ω is the time constant of integration (normally set to 2.5 ms), and $T = i\,dt$.

(7) averaging of ACFs across the 128 channels to produce a summary ACF,

$$s(t,\delta t) = \sum_{k=1}^{128} \frac{h(t,\delta t,k)}{128}. \quad (2)$$

(8) *either* (a) extraction of pitch by inspection of the summary ACF major peaks. We take the highest peak of the summary ACF within the pitch region between 60 and 400 Hz, *or* (b) estimation of discriminability of two stimuli by computing distance measures comparing the summary ACFs from the two stimuli:

$$D_t^2 = \sum_{i=1}^{L} \frac{[s(t,i\,dt) - s'(t,i\,dt)]^2}{L}, \quad (3)$$

where $s(t, i\,dt)$ and $s'(t, i\,dt)$ are the summary ACFs at time t of the two stimuli, $i\,dt$ is the autocorrelation lag, L is the number of autocorrelation lags used (normally 334); when L is 334, the ACF includes pitches down to pitches of 60 Hz for a sample rate of 20 kHz; D_t^2 is a function of time and, even for steady-state stimuli, is subject to small oscillations when the time constant Ω is small relative to the period of the signal.

This problem is unavoidable because the similarity of two signals will necessarily vary along the length of the signals. We shall return to the problem later when discussing mistuned harmonics. We routinely measured the value of D_t^2 at the end of the stimulus which was always a multiple of the signal period except where explicitly indicated.

We have no direct measure of the threshold of discriminability but assume that the threshold is lower when D^2 is large. Accordingly, we have used the $1/D^2$ as our measure of the "relative threshold." This does not tell us exactly where the threshold is but does allow us to predict the directional effect of stimulus manipulations on the threshold.

The companion article (Meddis and Hewitt, 1991) should be consulted for detailed examples of the operation of the model at each stage. The results in this paper will be given entirely in terms of the summary ACFs which are the main output of the model in response to a stimulus.

II. EVALUATION OF THE MODEL

A. Wightman's counter example

Wightman (1973a) argued against the "peak-picker" or "fine-structure" theories of pitch such as the theory of Bilsen and Ritsma (1969) by saying that their pitch predictions appeared to be phase-sensitive. Psychoacoustic experimentation (e.g., Patterson, 1973) had previously shown that pitch matching was largely unaffected by the relative phases of the stimulus tone components. As a part of his argument, Wightman demonstrated that listeners showed no change in pitch percept using a stimulus which, according to peak-picker theories, ought to have produced a substantial change. We shall begin by examining this demonstration because it has been very influential and is uppermost in the minds of many researchers when attempts are made to reintroduce temporal theories into the debate.

Figure 2 is a 1-kHz carrier tone, 100% amplitude modulated with a 200-Hz sinusoid. Below it is the same function but inverted. Wightman(1973a) reported that listeners could not distinguish between the two signals. A "peak-picker" model would find only one prominent peak per cycle in the first wave but two prominent peaks per cycle in the second wave. Any pitch prediction involving time intervals between the prominent positive peaks of the input signal should predict different perceived pitches for the two stimuli. This kind of peak-peaking analysis was, therefore, contradicted by the fact that subjects heard no difference.

The model proposed here also depends on time intervals. However, the analysis is based on the individual bandpass filtered signals found in the auditory nerve and not on the original signal. This distinguishes the model from the traditional peak-picking approach.

An amplitude-modulated signal of this type can be decomposed into the three harmonic spectral components:

$$x(t) = 0.5m \cos[2\pi(n-1)ft\,] + \cos[2\pi nft\,] + 0.5m \cos[2\pi(n+1)ft\,], \quad (4)$$

where f is the fundamental frequency which also represents the true period of the waveform, n is integer and m is the modulation index. For this signal the three spectral components are (800, 1000, and 1200 Hz) and they are almost completely resolved by peripheral (cochlear) frequency analysis. Accordingly, our autocorrelation analyses are applied to three separate tones occupying their own channels.

Figure 2 shows the output of the model when presented with this stimulus in both regular and inverted form. The summary ACFs for these two waveforms[2] are almost identical implying that listeners would not be able to distinguish between the two signals. Because the system "sees" the signal as three almost wholly resolved sinusoidal components, the major time intervals are the intervals between peaks of the individual sinusoids. These are unchanged when the stimulus waveform is inverted.

The value of this test is that it distances our model from a simple peak-picker interpretation. Criticisms of the latter do not automatically apply to our model. Below, it shall be shown that the model is, indeed, sensitive to phase under certain circumstances but in this particularly famous case it is not.

B. Quasifrequency modulation (QFM)

Ritsma and Engel (1964) sought to demonstrate that changes to the fine structure of the stimulus waveform would influence the virtual pitch of the stimulus. They used a quasifrequency-modulated (QFM) signal consisting of only three frequency components where the phase of the center component is shifted by 90 deg:

$$x(t) = 0.5m \sin[2\pi(n-1)ft\,] + \sin[2\pi nft + \pi/2] + 0.5m \sin[2\pi(n+1)ft\,]. \quad (5)$$

Ritsma and Engel held nf (the carrier frequency) constant at 2000 Hz.

While $1/f$ is the true period of the signal, Ritsma and Engel showed that QFM signals of this type have a pseudo-period of approximately $1/2f$, which is especially clear if the modulation index m, is greater than 1. They claimed that their subjects could identify pitches at around both the fundamental frequency and a frequency an octave above (corresponding to the pseudo-period of the waveform). For n equal to 11 and 13, subjects matched to pitches very close to f and $2f$, while for n equal to 10 and 12 they matched to pitches closely *above* or closely *below* f and $2f$, but rarely between. They explained their results in terms of the time intervals between the major individual peaks of the stimulus waveform.

This was an important result because it appeared to represent a clear confirmation of the "peak-picker" approach. This predicted a different result to an analysis of time intervals between the peaks of the *envelope* of the signal that predicted pitch matches at f and $2f$. However, Wightman

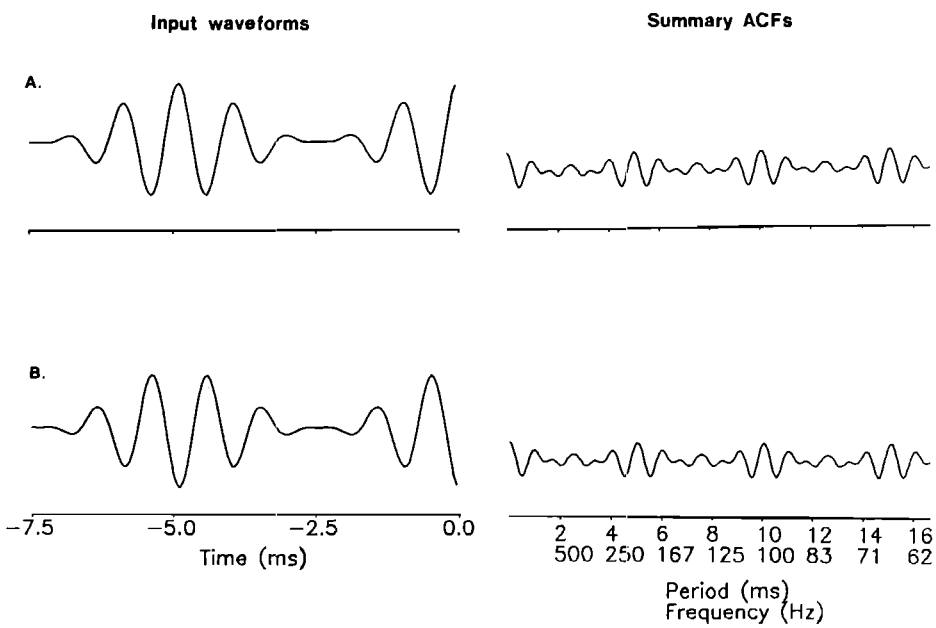

FIG. 2. Wightman's peak-picker counterexample. The first stimulus consists of a 1-kHz carrier tone 100% amplitude modulated with a 200-Hz sinusoid; the second is an inverted version of the first. The summary ACFs for both are identical indicating that the model does not discriminate between them (see text).

(1973a) replicated the experiment and found only pitch matches in the region of the fundamental frequency (f) with no ambiguous pitches on either side. This refutation of Ritsma and Engel's result has been influential in establishing the view that the human pitch detection mechanism is phase insensitive. We, therefore, decided to present these QFM signals to our model using $n = 11$ and $n = 12$ with $nf = 2000$ Hz and $m = 2.55$. The results are given in Fig. 3.

The summary ACFs show clear peaks at the fundamental frequency (f) for both odd and even values of n and, therefore, agree with Wightman's observations and disagree with those of Ritsma and Engel. The model does not predict pitch matches close to but above and below f. However, the summary ACFs also show major peaks near to twice the fundamental frequency; for n odd the peak is exactly located at $2f$ [Fig. 3(b)], while, for n even, two peaks are located above and below $2f$ [Fig. 3(a)]. This latter result agrees with Ritsma and Engel's observation. Unfortunately, Wightman did not report looking for pitches in this region an octave above f; he was content to refute the suggestion that matches could not be found close to but different from f. In conclusion, the model results agree with Wightman's observations at the fundamental frequency but also agree with Ritsma and Engel's observations at the octave. They are also consonant with results reported in Moore's (1977) study of phase inversion.

C. Alternating phase

Patterson (1987) systematically explored the listener's ability to discriminate between two harmonic tone complexes which differed only in terms of the phase relationships among their pure-tone components. He shifted the phase of alternate harmonics in the alternating phase (APh) complex:

$$x(t) = a \sum_{n=h}^{h+k} \cos(2\pi fnt + \phi_n),$$

$$\phi_n = 0, \text{ for } n \text{ even}, \quad \phi_n = j, \text{ for } n \text{ odd}, \quad (6)$$

where k is the number of harmonics, h is the number of the lowest harmonic (called the "harmonic number" below), and j is the phase shift for alternate harmonics. By shifting the phase for every alternate harmonic, Patterson was able to optimize the conditions for hearing phase effects. With appropriate fundamental frequencies and given adequate numbers of harmonics, there is a very clear contrast with a harmonic stimulus with all components in cosine phase (CPh). Using this teminology, a CPh signal has ϕ_n set to zero for all values of n.

Patterson measured the minimum detectable phase shift j, when an APh signal was compared with a CPh stimulus. He found that thresholds varied as a function of the number of the lowest harmonic (h), the frequency of the fundamental (f), and the level of the stimulus (a). The duration of the stimulus (between 32 and 512 ms), however, did not affect the threshold for the detection of phase alternation. Below, we compare the predictions of our model with Patterson's empirical results.

As an introductory example, Fig. 4 illustrates the summary ACFs for a 20-harmonic tone complex with alternating phase (APh) and cosine phase (CPh). In the APh stimulus, the phase of odd-numbered harmonics has been advanced 30 deg with respect to the even-numbered harmonics. Shifting the phase of alternating harmonics produces a notch in the stimulus waveform at the midcycle point. The summary ACFs produced by the model to both APh and CPh stimuli are superimposed in Fig. 4, where it can be seen that there are small differences. At the bottom of Fig. 4 the differences between the summary ACFs are shown in magnified form. These differences are combined using the Euclidean distance (D^2) measure given above [Eq. (3)]. The magnitude of the distance indicates the model's prediction of the extent to which listeners can discriminate between the two.

The differences between the two summary ACFs arise from small differences in the ACFs for the individual frequency-selective channels. Phase changes are therefore best detected by the model when there is harmonic interaction

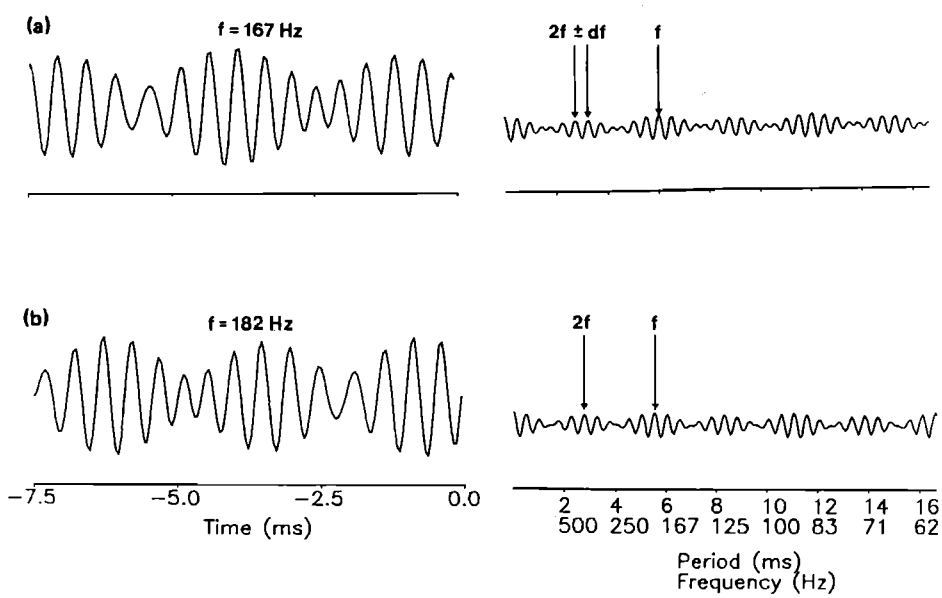

FIG. 3. Quasifrequency modulation. The stimuli are two QFM signals with 2-kHz carrier tone and modulation rates of (a) 166.7 Hz and (b) 181.8 Hz. The modulation depth (m) is 2.55 in both cases. The summary ACFs both predict pitch matches at exactly the fundamental but also predict matches near twice the fundamental frequency. In (a) matches are predicted slightly higher and slightly lower than $2f$, while in (b) a pitch match is predicted at exactly $2f$.

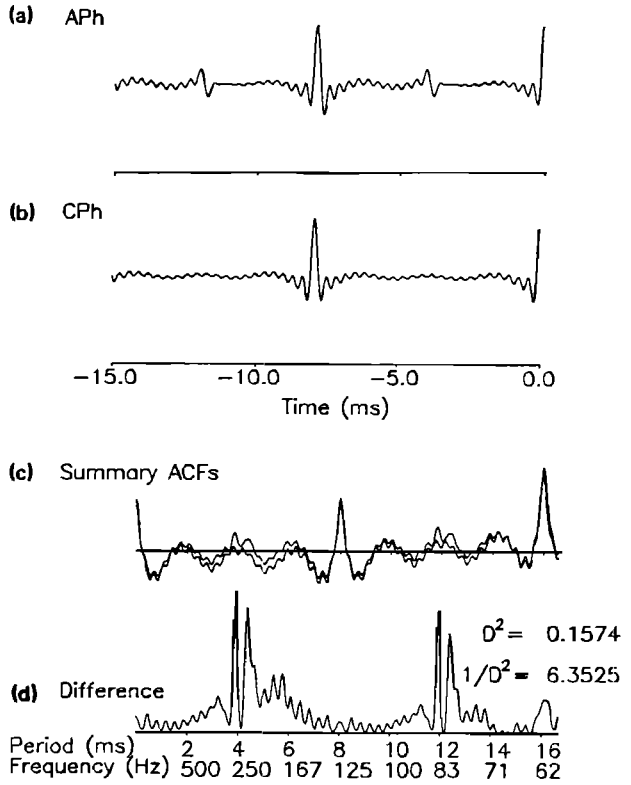

FIG. 4. Alternating phase (APh). The two stimuli consist of 20 harmonics of a 125-Hz tone. In (a) odd harmonics are 30° out of phase (APh). In (b) the components are in cosine phase (CPh). (c) The summary ACFs for the two stimuli, superimposed for comparison purposes and (d) the squared differences between corresponding parts of the two summary ACFs. The last 15 ms of the signal are shown for clarity but the time constant of the model remains unchanged at 2.5 ms. The ACFs are computed 128 ms after stimulus onset.

within channels. Note that the stimuli in Fig. 4 are close to the threshold of discriminability for listeners.

To recapitulate, the procedure for testing the model involves comparing the responses of the model to a cosine phase (CPh) stimulus and to an alternating phase (APh) signal. A summary ACF is generated for each and compared along its length using the Euclidean distance measure given above [Eq. (3)]. The graphs show the *reciprocal* of the Euclidean distance as an indication of the "relative threshold" of the stimulus. Summary ACFs, which are clearly discriminable, have large Euclidean distances and small reciprocals (low relative thresholds). This form of presentation allows a rough comparison with the *trends* present in Patterson's results. We shall be concerned with the sensitivity to phase shifts as a function of (a) fundamental frequency, (b) lowest harmonic number, (c) signal amplitude, and (d) signal duration. In what follows, we have used the model that was developed for simulating pitch perception without any variation whatsoever. Suggestions for varying parameters to improve the results will be given after the presentation of the original results.

Except where a stimulus parameter is explicitly varied, the stimuli consist of 20 harmonics from the 4th to the 24th harmonic and are presented at 50 dB1 (see footnote 3). The summary ACF is read 128 ms after stimulus onset. The sample rate of the model is 20 kHz. This rate was chosen because the results were insensitive to sample rate above 20 kHz. All APh waves used in the testing of the model have a phase shift of 40 deg, a value that Patterson's results show to be near threshold for many of his stimuli.

1. Harmonic number and fundamental frequency

Figure 5 shows the effect of varying the number of the lowest harmonic and the fundamental frequency for stimuli with only eight harmonics. Patterson found that thresholds were lower for stimuli with lower fundamental frequencies and this is clearly reflected in the results of the model. This effect may be attributed to harmonic interaction within frequency-selective channels. Lower fundamental frequencies give rise to harmonics that are more closely spaced and that will produce more between-harmonic interaction within channels. It is this interaction that allows the phase relationships between adjacent harmonics to be detected on a within-channel basis.

Patterson also noted that threshold falls as the harmonic number is increased, a result that the model simulates moderately successfully. For 250-Hz fundamental stimuli the trend is quite clear. However, for 62.5- and 125-Hz stimuli the decrease levels off after eight harmonics. While the model simulates the empirical data tolerably well, it does leave room for improvement and we shall consider this below.

Again, we can attribute the reduced thresholds to harmonic interaction within channels. The bandwidth of the filters is greater at higher center frequencies and the interaction of adjacent harmonics is greater as a consequence when the harmonic number increases.

2. Amplitude

The effect of *amplitude* is shown in Fig. 6. There is a gradual decline in relative threshold as amplitude is increased. The suggested explanation for this effect is as follows. As level increases, the firing rate of the individual fibers rises causing an increased number of interspike intervals in the summary ACFs. This serves to exaggerate the Euclidean distance between the summary ACFs of the two stimuli. Note that neither the ACF computation nor the Euclidean distance measure is corrected in any way for overall activity level.

Patterson's data show little or no decline in threshold for the 250-Hz fundamental stimulus as amplitude increases. Our data show a clear decline. We have no explanation for this except to point out that his subjects did have considerable difficulty in detecting phase changes for these stimuli. Two of his listeners were unable to perform the test reliably and were omitted. His figures are, therefore, based on only the two remaining subjects and need not be strictly comparable with the functions for the other two fundamental frequencies.

The model output above 50 dB1 shows a floor effect. This is caused by the limited dynamic range of the hair-cell model, which shows little increase in the rate of firing above 50 dB1. We expect this effect to be reduced, somewhat, in an extended simulation which included a range of types of AN

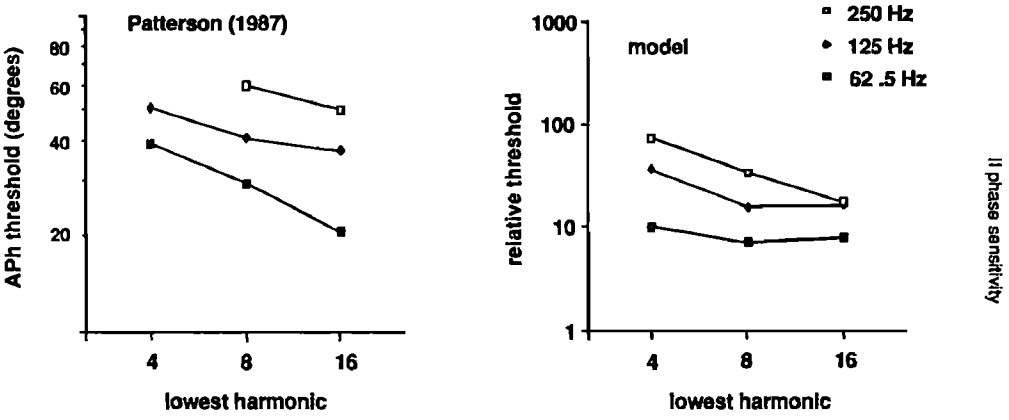

FIG. 5. Lowest harmonic number (APh). Relative threshold (1/Euclidean distance) for eight equal-amplitude harmonic stimuli in alternating phase for three fundamental frequencies. All stimuli have alternating components that are 40° out of phase, last for 128 ms, and are presented to the model at 50 dB1 per component.

fibers and we shall address the issue below. (The significance of this discrepancy is somewhat complicated by the fact that his data *do* show a floor effect in response to greater amplitude with monotonic phase stimuli—see Fig. 11, below.) However, a complete explanation for the discrepancy may lie with the linear nature of the filters used in this implementation. At high stimulus levels, the peripheral filtering of the basilar membrane assumes broader bandpass characteristics (e.g., Robles *et al.*, 1986), which are detectable in psychophysical experiments with human listeners (Patterson and Moore, 1986). These changes are attributable to a nonlinear response of the basilar membrane to intense stimulation. Unfortunately, the filters we have used are linear. If we had used nonlinear filters which showed a wider tuning curve at higher intensities, this would have increased the number of harmonic components that could interact within a given channel and thus enhanced the response to phase effects. Since the effects of nonlinearity do not appear below about 40 dB SPL, we would expect the nonlinear effect of level to contribute maximally above 50 dB SPL and remove the floor effect visible in Fig. 6 (see footnote 4).

3. Duration

The effect of varying the *duration* of the stimulus is shown in Fig. 7. For this test we sampled the summary ACF after 32, 64, 128, and 256 ms. We were unable to include 512-ms durations because of the storage limitations of our computer implementation, but there is no reason to expect that the result would be substantially different from that for 256 ms. The model results agree with Patterson's data in that they show very little change with time.

This is not a very surprising result as far as the model is concerned because it has a memory-less operation: It contains no mechanism whereby it could accumulate information across the full duration of the stimulus. What is surprising is Patterson's discovery that human listener's operate in this way. In almost all other psychophysical experiments, the nervous system does seem to benefit from longer (up to 500 ms) exposure to a stimulus.

D. Monotonic phase change across harmonic components

Patterson also studied sensitivity to phase changes that were applied smoothly across the spectrum. The phase of each successive harmonic was advanced by a small amount with respect to the phase of the previous harmonic to produce a monotonically increasing phase spectrum such as that given in Fig. 8. This monotonic phase (MPh) waveform was created so as to minimize phase changes within auditory channels but to maximize phase changes across channels:

$$x(t) = a \sum_{n=h}^{h+k} \cos(2\pi f n t + \phi_n), \tag{6}$$

where ϕ_n, the phase change, was calculated to give a fixed

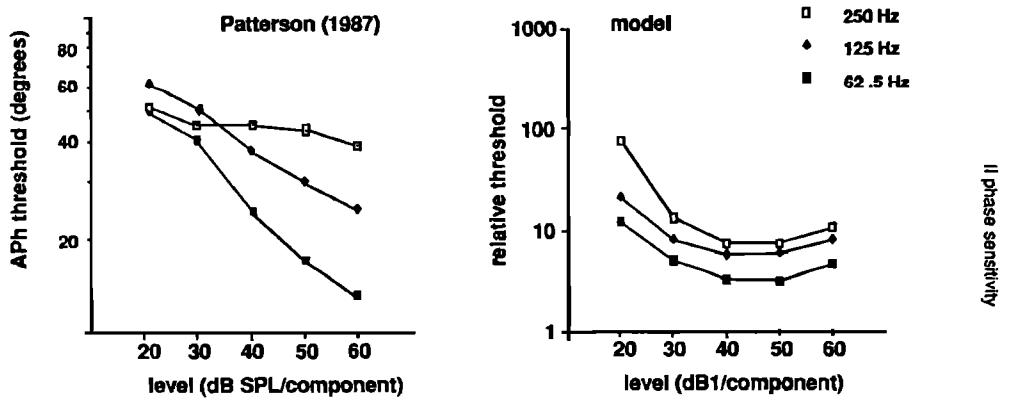

FIG. 6. Amplitude (APh). Relative threshold (1/Euclidean distance) for 20 equal-amplitude harmonic stimuli in alternating phase for three fundamental frequencies. All stimuli have alternating components which are 40° out of phase, last for 128 ms, and have a lowest harmonic number of 4.

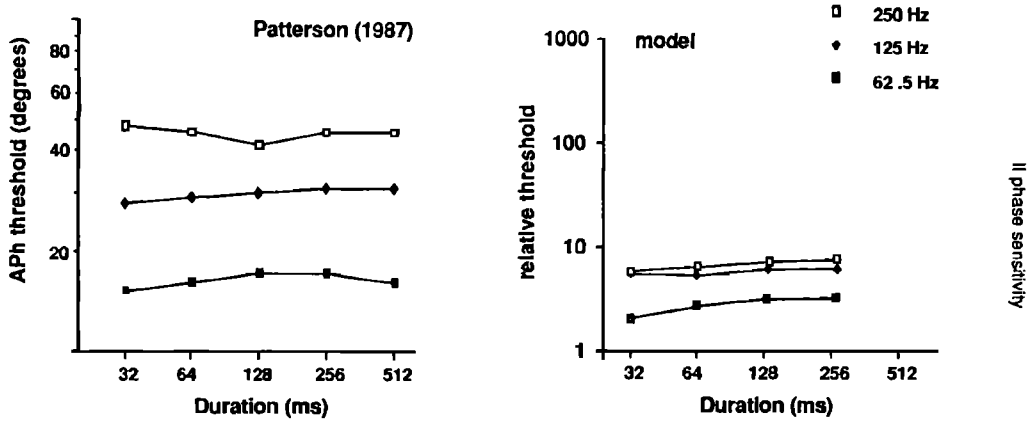

FIG. 7. Duration (APh). Relative threshold (1/Euclidean distance) for 20 equal-amplitude harmonic stimuli in alternating phase for three fundamental frequencies. All stimuli have alternating components that are 40° out of phase, have a lowest harmonic number of four, and are presented at 50 dB1 per component.

amount of phase shift between the lower and upper 3-dB points of each channel filter. The rate of change of phase across a given channel is called the "slant." The method of calculation and some sample values are given in Patterson (1987). Figure 8 shows an example of the relative phases of the successive harmonics of a 128-Hz MPh wave for a slant of 1. A positive slant has the effect of advancing low frequencies relative to high frequencies.

As before, the subject's task was to discriminate between two stimuli: a CPh version of the waveform and an MPh version. Patterson's (1987) data typically indicate threshold performance for slant values in the region of unity. Accordingly, when testing the model, we have fixed the slant at 1 and measured the Euclidean distance between the summary ACF for the CPh and MPh waveforms. We then used the reciprocal of the distance as a measure of relative threshold. The phase of any component can therefore be found using Fig. 8.

The stimulus parameters (level, duration, number of harmonics, number of lowest harmonic) were the same as those used in the APh study reported above. Figure 9 illustrates the testing of the model using an MPh stimulus consisting of the 4th to 24th harmonics of 125-Hz fundamental with a phase slant of unity.

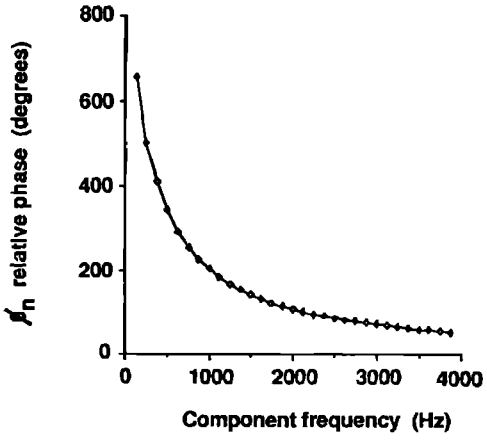

FIG. 8. Phase spectrum for MPh (monotonic phase) stimuli. The relative phase is given in degrees for a "slant" of 1. For other slants, the relative phase is adjusted by multiplying by the new slant. The phase spectrum is designed to minimize the amount of within-channel phase change.

1. Harmonic number and fundamental frequency

Figure 10 illustrates the effect of varying the number of the lowest harmonic and fundamental frequency in an eight-harmonic tone complex. Patterson found that there was no measurable superiority at all for low fundamental frequen-

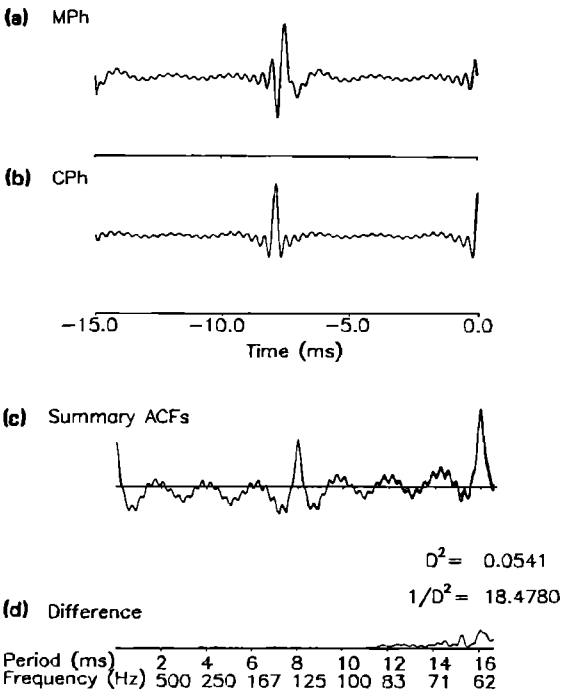

FIG. 9. Monotonic phase (MPh) wave with a slant of 1. The two stimuli consist of 20 harmonics of a 125-Hz tone. In (a) the phase of successive components are advanced with a slant of 1 (see Fig. 8). In (b) the components are in cosine phase (CPh). (c) The summary ACFs for the two stimuli, superimposed for comparison purposes and (d) the squared differences between corresponding parts of the two summary ACFs. The last 15 ms of the signal is shown for clarity but the time constant of the model remains unchanged at 2.5 ms.

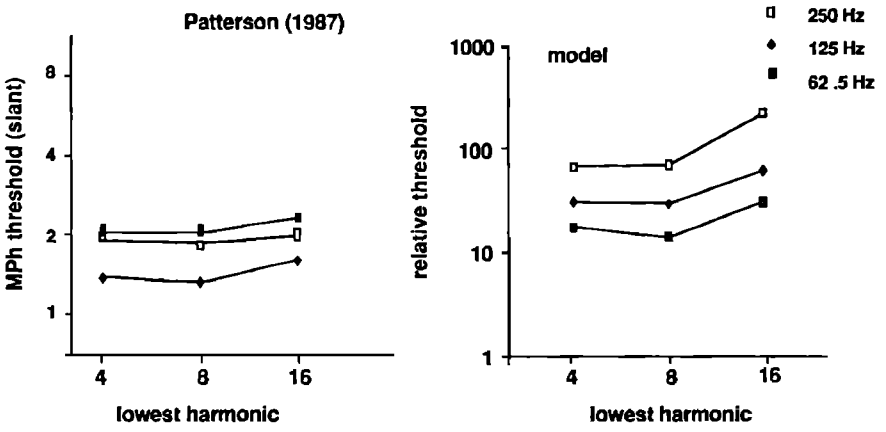

FIG. 10. Lowest harmonic number (MPh). Relative threshold (1/Euclidean distance) for eight equal-amplitude harmonic stimuli in monotonic phase for three fundamental frequencies. The phase of successive components are advanced with a slant of 1 (see Fig. 8), the stimuli last for 128 ms, and are presented to the model at 50 dB1 per component.

cies with MPh stimuli unlike APh stimuli. It was even likely that low fundamental frequencies were associated with slightly higher thresholds. The model, however, continues to show lower thresholds for low fundamental frequencies. It does so for the same reason as given for APh stimuli; the harmonics are more closely packed and there is more opportunity for within-channel interaction between harmonic components of the stimuli. This result shows a clear and unexplained discrepancy between the model and the human listener data.

The effect of varying the lowest harmonic number, however, shows a clear parallel with Patterson's results. Whereas for APh stimuli, thresholds fell with higher harmonic number, both model and human data agree that varying the number of the lowest harmonic had very little effect. If anything, the model showed an increase in threshold as the harmonic number was raised from 8 to 16. The same tendency can be seen in Patterson's data.

The lower thresholds for higher harmonic numbers was attributed, in the case of APh, to wider filter bandwidths at higher frequencies, while harmonic spacing remained constant. In the case of MPh stimuli, the phase shift between adjacent harmonics is *reduced* at higher frequencies. The rate of reduction of phase difference is linked to the rate of increase of the filter bandwidths in such a way (see Fig. 8) that the two effects should cancel. As a consequence, the harmonic number effect found for APh is not present for MPh.

2. Amplitude

The effect of *level* is approximately the same as for APh waveforms, that is, a decline in threshold with stimulus level (Fig. 11). The model results agree broadly with Patterson's data in this respect. On this occasion, Patterson's data show a floor effect above 40 dB, while the model does not. The model did show a floor effect for the corresponding APh stimulus, however.

3. Duration

Our results show no change in threshold as a function of *stimulus duration* between 64 and 256 ms (Fig. 12). This is the same as for APh waveforms and in full agreement with Patterson's results. However, the threshold for the model is lower at 32 ms. This is caused by the adaptation of the simulated auditory-nerve fibers in the first 50 ms of the stimulus. The firing rate at stimulus onset is accompanied by better discrimination. It follows that the model would give the best discrimination if the judgment were made on the basis of the first 32 ms of the stimulus and the rest of the stimulus were ignored. The dotted lines in Fig. 12 indicate the expected performance if this stratagem were to be employed. The resulting parallel functions reflect Patterson's results and offer an explanation of why his subjects did not find the task any easier as the stimulus duration increased.

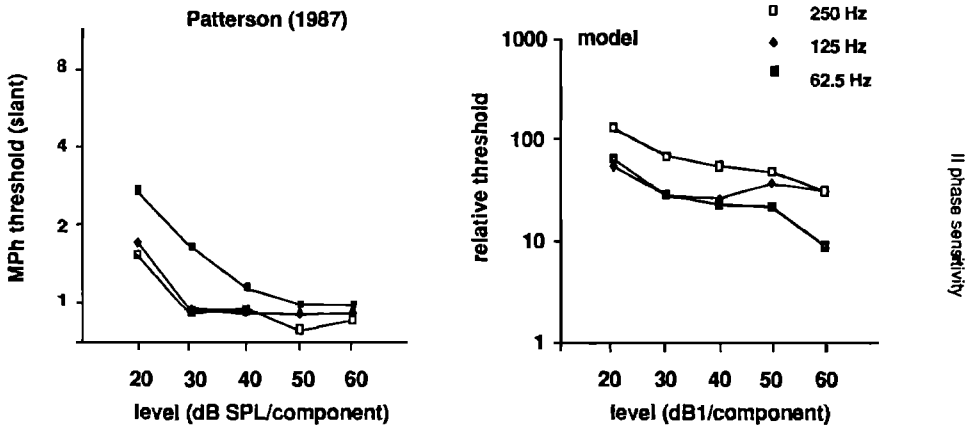

FIG. 11. Amplitude (MPh). Relative threshold (1/Euclidean distance) for 20 equal-amplitude harmonic stimuli in monotonic phase for three fundamental frequencies. The phase of successive components are advanced with a slant of 1 (see Fig. 8), the stimuli last for 128 ms, and have a lowest harmonic number of 4.

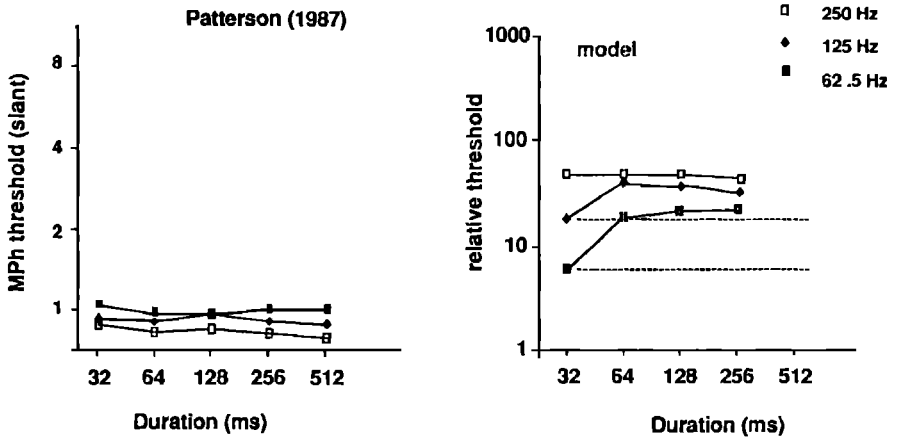

FIG. 12. Duration (MPh). Relative threshold (1/Euclidean distance) for 20 equal-amplitude harmonic stimuli in monotonic phase for three fundamental frequencies. The phase of successive components are advanced with a slant of 1 (see Fig. 8), the stimuli have a lowest harmonic number of 4, and are presented at 50 dB1 per component.

4. Summary of MPh and APh evaluations

There is considerable broad agreement between the model performance and Patterson's data for human listeners. It is difficult to know how much significance to attach to some of the more subtle differences because we have no indication of the reliability of the phenomena. Patterson (1987) used only four subjects and some of these subjects were not able to perform all of the tests. Moreover, no indication of the variation in individual performance was given in the article.

Two discrepancies did give rise to concern, however. Both are illustrated in Figs. 6 and 10, where lower fundamental frequencies give rise to lower thresholds for APh, while the opposite is the case for MPh. We evaluated the model a second time using a much longer time constant ($\Omega = 20$ ms) because we were concerned that the time constant was very short relative to the period of the stimuli. In general, this change had little effect on the results and only the harmonic number study is reported here for both APh and MPh (Fig. 13). While the APh result showed a more realistic fall in threshold across harmonic number, the change in time constant had no effect on the relationship with the fundamental. The discrepancy with Patterson's results remains unexplained.

E. Mistuned harmonics

Hartmann (1988) has reported a remarkable demonstration of listeners' sensitivity to the phase of a single component of a harmonic complex. By employing stimuli of varying duration, he was able to plot plateaux and troughs in discrimination ability as a mistuned harmonic goes in and out of phase with a harmonic complex. He used 40-dB harmonic complexes composed of the first seven harmonics of a fundamental in the region of 200 Hz as his comparison stimulus. The test stimulus was the same except for the fourth harmonic, which was mistuned by 2.5% (e.g., from 800 to 820 Hz). For long duration stimuli (> 100 ms), listeners were able to perform this discrimination at better than 90% levels of success. For shorter duration stimuli, performance improved gradually from around chance levels at 15-ms duration. However, the improvement was nonmonotonic with a distinct trough at around 50 ms [Fig. 14(a)].

He tested the hypothesis that this trough was a stimulus-phase effect by repeating the analysis with the mistuned harmonic 180 deg out of phase. The improvement in performance between the 15- and 90-ms stimulus durations was again nonmonotonic but the trough was now at 25 ms with a peak at 50 ms [Fig. 14(a)]. This result is a clear demonstration of phase sensitivity.

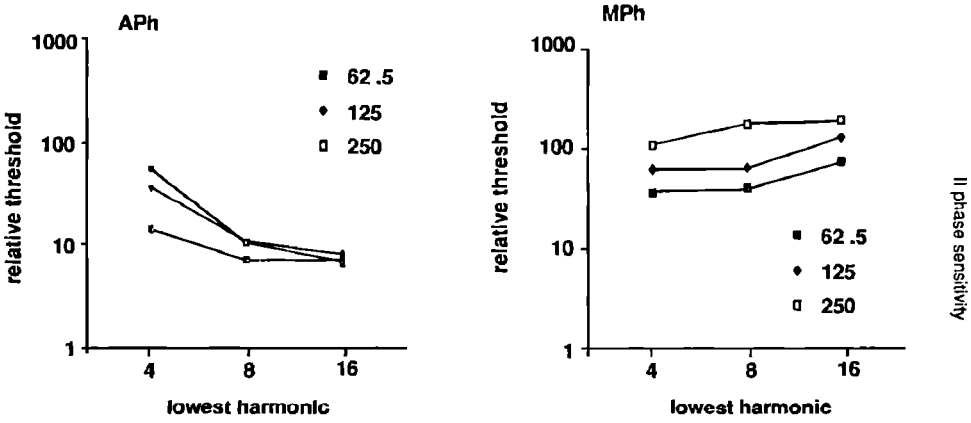

FIG. 13. The effect of using a 20-ms time constant. The corresponding figures for comparison are Figs. 5 and 10.

Time constant 20 ms

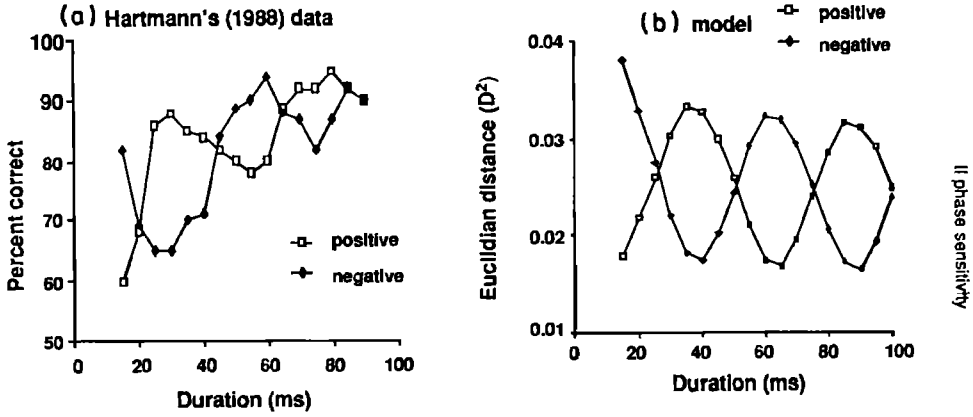

FIG. 14. The effect of a 2.5% mistuning of the fourth harmonic of a seven harmonic complex with a fundamental frequency of 200 Hz. Subjects were asked to identify the stimulus with the mistuned component in a comparison with a harmonic stimulus. Open squares: All components start at positive-going zero crossing. Closed diamonds: Mistuned harmonic begins 180 deg out of phase. (a) Data taken from Hartmann (1988). (b) Results of model evaluation for the same stimuli.

We tested the model using Hartmann's stimuli with a range of durations from 20 to 90 ms with increments of 5 ms. The Euclidean distance, D^2 [see Eq. (3) above] was calculated between a CPh stimulus and a similar stimulus with the fourth harmonic mistuned by 2.5%. D^2 was computed over a range of lags restricted to the "timbre region" ($0 < \delta t < 0.004$s). D^2 was used as a measure of discriminability that should predict percent correct answers at least in terms of directional trends. The results of this investigation are given in Fig. 14(b) and the sensitivity of the model to the relative phase of the mistuned harmonic is clear.

The model does not, however, show the underlying trend of gradual improvement with increasing duration because it has no mechanism, as yet, for aggregating information over long durations. The only integration that does take place is due to the time constant (2.5 ms) of the running ACF calculations, which is too short to be relevant to this problem.

III. DISCUSSION

Until recently, it was generally agreed that listeners were unable to discriminate changes in the phase spectra of monaural stimuli and models of auditory perception reflected this belief. Wightman (1973a) made a virtue of the phase insensitivity of his model and criticized temporal theories because they appeared able to discriminate between certain signals with different phase spectra when it was manifest that human listeners heard them as similar. It now appears that people are phase sensitive but can only discriminate certain kinds of phase changes. As a consequence, models are now required to reflect the new position. Temporal theories are obvious candidates but the early "peak-picker" theories have been discredited because they are sensitive to too many different kinds of phase change.

The model we propose has the virtue of predicting a wide range of psychophysical results in both pitch perception and phase sensitivity. It is a temporal model in the sense that the autocorrelation functions are equivalent to the aggregation of time intervals between spikes in the auditory nerve. It differs from peak-picker theories in that it performs analyses within frequency-selective channels and not directly on the stimulus waveform. As a consequence, the new model is mainly sensitive to *within-channel* phase spectra changes while peak-picker theories responded to all phase spectra changes.

The proposed model is not without its problems, however. The failure to predict that higher fundamental-frequency, MPh, harmonic complexes produce better sensitivity to phase change is an unresolved issue. Some modification to the model is clearly required but it is not clear what it might be. Alternatively, the problem may be caused by inaccurate modeling of filter widths. The phase characteristics of the stimuli were based on an *assumed* set of listener filter characteristics. Any weakness in these assumptions might lead to unpredictable consequences for performance.

Less seriously, the failure to replicate the underlying improvement in performance with stimulus duration in Hartmann's (1988) task clearly requires attention. Unfortunately, Patterson (1987) found no improvement in stimulus discriminability with stimulus duration and this complicates the modeling problem. Their stimuli were slightly different in that Hartmann used a steadily advancing phase difference while Patterson used fixed phase characteristics. The implications of these results for modeling are not yet clear.

The peripheral aspects of the model are firmly anchored in a generally agreed approach to the activity of the basilar membrane and the auditory nerve. There is, of course, room for improvement here by way of better specification of filter bandwidths, the development of reliable nonlinear models of cochlear filtering and a greater variety of fiber types (low-, medium-, and high-spontaneous rates) with different thresholds and rate-intensity functions. Nevertheless, the conceptual bedrock is reasonably firm.

By contrast, we know little about the responses of neurons in the auditory nervous system to phase spectra changes. In the companion article (Meddis and Hewitt, 1991), we speculate concerning the possible physiological basis for amplitude modulation sensitivity. Such speculation is not possible for phase data because the necessary investigations have not been attempted.

The relative merits of our model with respect to other theories have been discussed in the context of pitch perception in our previous article (Meddis and Hewitt, 1991). However, phase sensitivity adds a further dimension to the

discussion. If we take phase into account, we may immediately set aside spectral models such as the "optimal processor" model of Goldstein (1973), the "pattern transformation" model of Wightman (1973b), and the virtual pitch models of Terhardt (1974, 1979) because they explicitly rule out phase sensitivity. It may be that they can be reformulated to take phase into account but this would be a project for the future. In addition, Wightman's (1973a) critique of peak-picker theories (Bilsen and Ritsma, 1969) is still valid and rules out one of the most influential temporal theories, which we do not intend to reinstate.

This leaves a group of theories that all advocate a temporal analysis of the array of bandpass-filtered signals leaving the basilar membrane (Moore, 1982; van Noorden, 1982; Patterson, 1987). An autocorrelation model is a particular instantiation of this genre in seeking to demonstrate that auditory perception depends on the timing information present in the spike activity of the auditory nerve. The only substantial difference between these theories and our model is that we have structured the model to produce quantifiable predictions and arranged a computational framework suitable for evaluating the theory against a range of existing empirical results.

Patterson's (1987) "pulse ribbon" approach is pictorial and his predictions are qualitative, while our approach is numerical and quantitative. Furthermore, we have evaluated our model against a wider range of pitch and phase data. Nevertheless, we do not quarrel with his basic approach, which emphasizes within-channel information concerning the timing of auditory-nerve spikes. However, the two models do differ in many points of detail. For example, the autocorrelation approach can use spikes arriving at any time, while the pulse ribbon model uses only spikes at the crests of the stimulus waveforms. To minimize between-channel phase effects, his model requires a loosely specified "alignment mechanism," while the autocorrelation function is insensitive to between-channel effects without further adjustment. Finally, the pulse ribbon model is schematic in explaining how the information is aggregated across channels before making the discrimination between two stimuli.

IV. CONCLUSIONS

A computational model using autocorrelation of bandpass-filtered signals has been shown to be useful in predicting a wide range of results from psychophysical studies of pitch perception and phase sensitivity. The evaluation reestablishes temporal analysis of auditory-nerve interspike intervals as a viable approach to modeling low-level auditory processing. However, there is considerable room for improvement of the basic peripheral model in terms of the range of types of auditory-nerve fibers used, the use nonlinear cochlear filtering, and the specification of filter bandwidths. Finally, there is also a need to identify physiological systems with analogous properties to the model.

ACKNOWLEDGMENTS

We would particularly like to acknowledge our debt to Roy Patterson, Brian Moore, Quentin Summerfield, and their colleagues for their comments and advice during the development of the model. This work was supported by a grant from the Science and Engineering Research Council Image Interpretation Initiative.

[1] Much of the early development of the model was based on a similar set of filters supplied by Martin Cooke, Department of Computer Science, Sheffield, UK.

[2] See Meddis and Hewitt (1991) for a detailed illustration of the steps between the presentation of the stimulus and the production of the summary ACFs.

[3] Stimuli presented to the model are simply number sequences but the dB1 scale is intended to reflect levels similar to the SPL scale. dB1 values are measured with reference to a signal with a root-mean square of 1. On a dB1 scale the hair cell model shows a rate-intensity threshold of 15 dB1 and a dynamic range of 30 dB.

[4] We have confirmed that a nonlinear model would reduce this floor effect. However, a full description of the nonlinear model is beyond the scope of this article.

Bilsen, F. A., and Ritsma, R. J. (**1969**). "Repetition pitch and its implication for hearing theory," Acoustica **22**, 53–73.

Buunen, T. F., Festen, J. M., Bilsen, F. A., and van den Brink, G. (**1974**). "Phase effects in a three-component signal," J. Acoust. Soc. Am. **55**, 297–303.

Goldstein, J. L. (**1973**). "An optimum processor theory for the central formation of the pitch of complex tones," J. Acoust. Soc. Am. **54**, 1496–1515.

Hartmann, W. (**1988**). "Pitch perception and the segregation and integration of auditory entities," in *Auditory Function*, edited by W. E. Gall and W. M. Cowan (Wiley, New York).

Licklider, J. C. R. (**1951**). "A duplex theory of pitch perception," Experientia **7**, 128–133.

Lundeen, C., and Small, A. M. (**1984**). "The influence of temporal cues on the strength of periodicity pitches," J. Acoust. Soc. Am. **75**, 1578–1587.

Meddis, R. (**1986**). "Simulation of mechanical to neural transduction in the auditory receptor," J. Acoust. Soc. Am. **79**, 702–711.

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

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

Meddis, R., Hewitt, M. J., and Shackleton, T. M. (**1990**). "Implementation details of a computational model of the inner hair-cell/auditory-nerve synapse," J. Acoust. Soc. Am. **87**, 1813–1816.

Moore, B. C. J. (**1977**). "Effects of relative phase of the components on the pitch of three-component complex tones," in *Psychophysics and Physiology of Hearing*, edited by E. F. Evans and J. P. Wilson (Academic, New York).

Moore, B. C. J. (**1982**). *An Introduction to the Psychology of Hearing* (Academic, London), 2nd ed.

Noorden, L. van (**1982**). "Two channel pitch perception," in *Music, Mind and Brain*, edited by M. Clynes (Plenum, London).

Patterson, R. D. (**1973**). "The effects of relative phase and the number of components on residue pitch," J. Acoust. Soc. Am. **53**, 1565–1572.

Patterson, R. D. (**1987**). "A pulse ribbon model of monaural phase perception," J. Acoust. Soc. Am. **82**, 1560–1586.

Patterson, R. D., and Moore, B. C. J. (**1986**). "Auditory filters and excitation patterns as representations of frequency resolution," in *Frequency Selectivity*, edited by B. C. J. Moore (Academic, London).

Patterson, R. D., Nimmo-Smith, I., Holdsworth, J., and Rice, P. (**1988**). "Spiral vos final report, Part A: The auditory filterbank," Cambridge Electronic Design, Contract Rep. (Apu 2341).

Ritsma, R. J., and Engel, F. L. (**1964**). "Pitch of frequency-modulated signals," J. Acoust. Soc. Am. **36**, 1637–1644.

Robles, L., Ruggero, M. A., and Rich, N. C. (**1986**). "Basilar membrane mechanisms at the base of the chinchilla cochlea, I. input-output functions, timing curves and response phases," J. Acoust. Soc. Am. **80**, 1364–1347.

Schouten, J. F. (**1970**). "The residue revisited," in *Frequency Analysis and Periodicity Detection in Hearing*, edited by R. Plomp and G. F. Smoorenburg (Sijthoff, Leiden).

Sivaramakrishnan, S., Long, G. R., and Tubis, A. (1989). "Phase sensitivity to amplitude-modulated stimuli as a function of component spacing," J. Acoust. Soc. Am. Suppl. 1 **85**, S121.
Terhardt, E. (1974). "Pitch, consonance and harmony," J. Acoust. Soc. Am. **55**, 1061–1069.
Terhardt, E. (1979). "Calculating virtual pitch," Hear. Res. **1**, 155–182.
Wightman, F. L. (1973a). "Pitch and stimulus fine structure," J. Acoust. Soc. Am. **54**, 397–406.
Wightman, F. L. (1973b). "The pattern transformation model of pitch," J. Acoust. Soc. Am. **54**, 407–416.