

# A unitary model of pitch perception

Ray Meddis and Lowel O'Mard

*Department of Psychology, Essex University, Colchester CO4 3SQ, United Kingdom*

(Received 15 March 1996; revised 22 April 1997; accepted 12 June 1997)

A model of the mechanism of residue pitch perception is revisited. It is evaluated in the context of some new empirical results, and it is proposed that the model is able to reconcile a number of differing approaches in the history of theories of pitch perception. The model consists of four sequential processing stages: peripheral frequency selectivity, within-channel half-wave rectification and low-pass filtering, within-channel periodicity extraction, and cross-channel aggregation of the output. The pitch percept is represented by the aggregated periodicity function. Using autocorrelation as the periodicity extraction method and the summary autocorrelation function (SACF) as the method for representing pitch information, it is shown that the model can simulate new experimental results that show how the quality of the pitch percept is influenced by the resolvability of the harmonic components of the stimulus complex. These include: (i) the pitch of harmonic stimuli whose components alternate in phase; (ii) the increased frequency difference limen of tones consisting of higher harmonics; and (iii) the influence of a mistuned harmonic on the pitch of the complex as a function of its harmonic number. To accommodate these paradigms, it was necessary to compare stimuli along the length of the SACF rather than relying upon the highest peak alone. These new results demonstrate that the model responds differently to complexes consisting of low and high harmonics. As a consequence, it is not necessary to postulate two separate mechanisms to explain different pitch percepts associated with resolved and unresolved harmonics. © 1997 *Acoustical Society of America*. [S0001-4966(97)06509-0]

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

## INTRODUCTION

The study of virtual (or low) pitch perception has a long history of controversy. Theories have typically championed either spectral (e.g. Terhardt, 1979; Goldstein, 1973) or temporal (e.g., Schouten, 1970; Licklider, 1951, 1959) explanations. Recently Houtsma and Smurzynski (1990) have shown that the pitch of stimuli consisting of high-numbered harmonics may be qualitatively different from and weaker than stimuli consisting of low-numbered harmonics. Patterson (1987) has also shown that listeners are better at detecting component phase shifts in such complexes when the harmonic numbers are higher. These and other results suggest that there may be two mechanisms for extracting pitch, one for low harmonics and one for high harmonics. Houtsma and Smurzynski (1990) give a useful discussion of this issue. Since spectral theories typically emphasize results based on low-numbered harmonics and temporal theories typically emphasize studies using high-numbered harmonics, this might be interpreted to mean that a spectral mechanism generates pitch percepts for low-frequency stimuli while a temporal mechanism generates pitch percepts for high-frequency stimuli. Carlyon and Shackleton (1994) have explicitly endorsed this view. The following investigations show, however, that both types of result can be embraced by a single model. This development offers the possibility that the traditional opposition of spectral and temporal approaches might be resolved in a single system.

We recently presented a computational model of pitch perception based on autocorrelation of the estimated probabilities of firing in groups of auditory nerve fibers (Meddis and Hewitt, 1991a, b). Although autocorrelation had been

used before in this context (e.g., Licklider, 1959; Lyon, 1984), the novelty of our system lay in the use of a summary autocorrelation function (SACF) as the basis of the prediction of the pitch percepts of human listeners (see also Assmann and Summerfield, 1990). The SACF is the vector sum across all audible frequencies of the autocorrelation functions (ACF) of auditory nerve-fiber firing probabilities. The idea of summarizing temporal information across channels is a characteristic of a number of other "temporal" models (van Noorden, 1982; Moore, 1977; Patterson, 1987).

Meddis and Hewitt's (1991a) model can be thought of as a special case of a class of a more general model consisting of four stages (see Fig. 1);

- (1) peripheral (mechanical) band-pass filtering,
- (2) half-wave rectification and low-pass filtering,
- (3) within-channel periodicity extraction,
- (4) across-channel aggregation of periodicity estimates.

Pitch decision algorithms are then applied exclusively to the results of the cross-channel aggregation in stage 4.

Meddis and Hewitt (1991a, b) used autocorrelation as one possible way of approximating the within-channel periodicity extraction stage. It was shown that the model was able to predict listeners' performance in the following areas; pitch of the missing fundamental, ambiguous pitch, pitch shift of equally spaced inharmonic components, musical chords, repetition pitch, the pitch of interrupted noise, the existence region, and the dominance region for pitch. The model was also able to simulate a number of aspects of listeners' sensitivity to the phase relationships among adjacent harmonic components of tone complexes. This phase sensi-

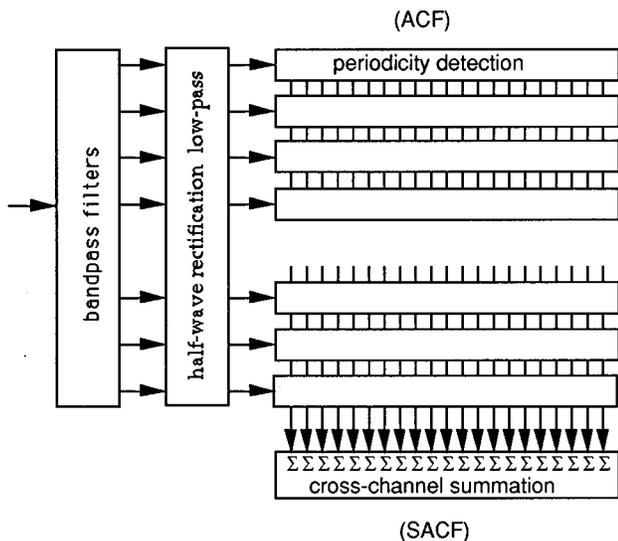


FIG. 1. General scheme of the four-stage pitch extraction model.

tivity was successfully demonstrated in the model using (i) amplitude-modulated and quasi-frequency-modulated stimuli, (ii) harmonic complexes with alternating phase and monotonic phase change across harmonic components, and (iii) phase effects associated with mistuned harmonics.

Such models have value in at least three different domains. First, they contribute to the development of psychological theories of the nature of pitch perception. Second, they serve as a class of computing engines acting as components of general models of auditory scene analysis (e.g., Brown and Cooke, 1994). Third, they offer a potential guide to physiological studies of the neurological mechanisms underlying pitch perception. Stages 1 and 2 of the model are consciously modeled on middle- and inner-ear functioning. Stages 3 and 4 are presumed to occur in the auditory brainstem although detailed theories of this process have yet to emerge. Our use of autocorrelation to simulate stage 3 of the model may appear unlikely as a physiological mechanism and the topic will be reviewed in the discussion.

In this report, we shall revisit the model by looking at three crucial phenomena not modeled before; (i) the pitch of alternating-phase harmonic complexes, (ii) frequency-difference limens for virtual pitch, and (iii) the effect of mistuning individual harmonics on the pitch of the whole complex. The phenomena are crucial in the sense that they could be used to justify dual (temporal/spectral) mechanisms of pitch perception. The presentation will emphasize an emergent property of the model whereby it responds to resolved and unresolved harmonics in a qualitatively different way without the need to postulate additional mechanisms.

## I. MODEL DESCRIPTION

Full details of the model are given in Meddis and Hewitt (1991a). It has not been altered in any substantial way. Some details of the implementation of the individual stages are given below and a list of all parameters is given in the appendix. The model was implemented using version 1.9.2 of the LUTEar computational library.<sup>1</sup> An example of the output of the model is given in Fig. 2 showing the ACFs and the

SACF in response to a 100-Hz fundamental tone consisting of groups of low- and high-numbered harmonics. The model was evaluated using a sampling interval of 50  $\mu$ s. All parameters are summarized in the Appendix.

*Stimulus generation.* The stimuli used were approximately 100 ms long with an onset (raised sine) ramp of 2.5 ms. The stimulus duration was always adjusted to be an integer multiple of the period of the signal. This is because the SACF pulsates in synchrony with the stimulus and the snapshot is best taken at a fixed point in the cycle. All tone components were harmonic and presented at 60 dB SPL unless otherwise stated.

*Pre-emphasis filter.* This filter is a simplified representation of the pressure gain of the middle ear. It is implemented using a second-order band-pass filter with skirts down by 3 dB at 450 and 8500 Hz.

*Band-pass filters (stage 1).* A bank of 60 linear fourth-order gammatone filters (Patterson *et al.*, 1992) was used to represent the mechanical frequency selectivity of the cochlea. Filter center frequencies were equally spaced on a log scale between 100 and 8000 Hz. The bandwidths of the filters were chosen using the formula suggested by Glasberg and Moore (1990).

*Hair cell (stage 2).* The output from each filter was passed to an inner hair cell model (Meddis, 1986, 1988), representing mechanical to neural transduction. [The parameters are the same as those given in Meddis *et al.* (1990) except for parameters *A* and *B* which were changed to 100 and 6000, respectively, to accommodate a change from arbitrary units to units of pressure,  $\mu$ Pascals.] The output from the simulated hair cell was in the form of a continuously changing probability of an action potential,  $p(t)$ , in a post-synaptic auditory nerve fiber. The output from the hair cell stage is characterised by an ac and a dc component. The ac component has the same frequency as the signal but is increasingly attenuated above 1 kHz. The dc component is a saturating monotonic function of signal level. Refractory effects were not modeled.

*Periodicity detection (stage 3).* A running autocorrelation function was computed in each channel based on the firing probabilities. The reported SACF is a "snapshot" of the function at the time of the cessation of the stimulus. More strictly, the function is the running cross product of the probabilities at a lag  $l$ ;

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

where  $h(t, l, k)$  is the ACF at lag  $l$ , time  $t$ , in channel  $k$ ,  $T = idt$ ,  $\tau$  is the time constant (normally 10 ms), and  $dt$  is the sampling interval.

*Cross-channel summation (stage 4).* All ACFs were summed with equal weight to create the SACF  $s(t, l)$ ,

$$s(t, l) = \sum_{k=1}^N h(t, l, k), \quad (2)$$

where  $N$  is the number of channels used.

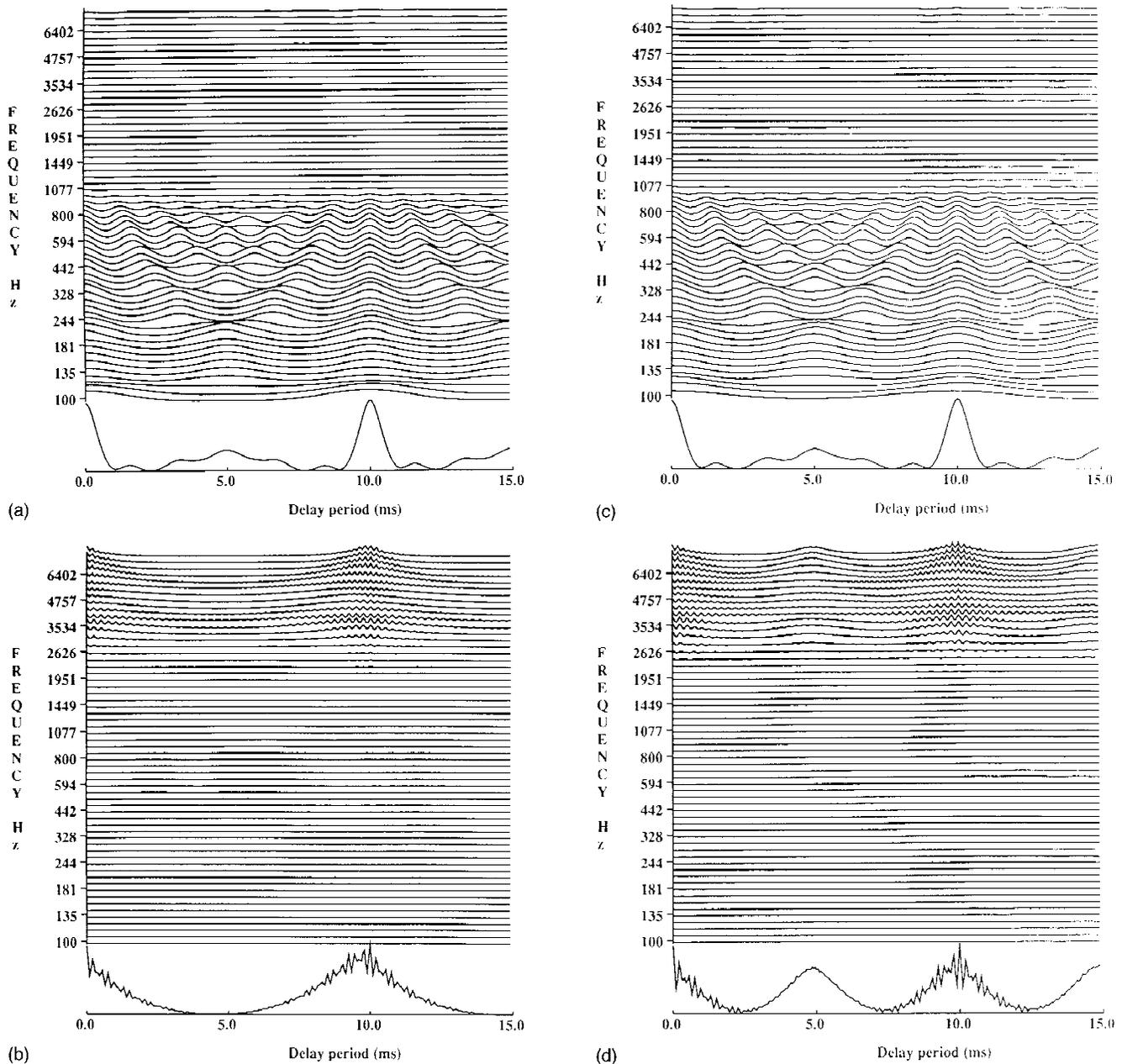


FIG. 2. Autocorrelation functions (ACF) and summary autocorrelation function (SACF) for harmonic complexes with 100-Hz fundamental. (a) and (b) Contrast results obtained using LOW (125–625 Hz) and HIGH (3900–5400 Hz) harmonics in *sine* phase. (c) and (d) Contrast results obtained using LOW and HIGH harmonics in *alternating* phase.

*Pitch matching* is implemented by generating the SACF for two tones and computing the squared Euclidean distance,  $D^2$ , between the two SACFs ( $s(t, l)$  and  $s'(t, l)$ ).

$$D^2 = \sum_{l=a}^b [s(t, l) - s'(t, l)]^2, \quad (3)$$

where  $l = i \cdot dt$ , and  $a, b$  defines the range of lags used in the comparison. Full details of this procedure are given in Meddis and Hewitt (1991b). A small  $D^2$  indicates that the two stimuli are similar while a large  $D^2$  indicates dissimilarity.

$D^2$  is measured over a limited segment of the SACF ( $a \cdot dt, b \cdot dt$ ) and the limits of this segment are explicitly specified on each relevant occasion in the text. The segment typically covers only those lags which might reasonably in-

fluence the pitch judgments, i.e., whose reciprocals represent frequencies at which virtual pitch can be heard. As a general rule, the *qualitative* effects to be described can be observed using a wide range of different segment lengths, so long as they include lags corresponding to the pitch period. To verify this effect, the model is routinely checked by using a different SACF ranges.

*Threshold judgments* are simulated in a *relative* sense by establishing  $D^2$  between a test stimulus and a comparison stimulus that differs by a fixed percentage along the dimension of interest (fundamental frequency or frequency of a mistuned harmonic). If the distance measure is greater in one situation than another then the former is judged to be more discriminable and to indicate a lower threshold (see below).

The test and comparison stimuli are chosen as typical of a threshold difference for human listeners. An alternative method used below is to define a reasonable criterion value of  $D^2$  and then to vary the fundamental frequency until this is matched. Both methods are illustrated.

In the case of *pitch matching*,  $D^2$  is computed between the test stimulus and a number of candidate comparison stimuli to be described below. The best match is taken to be the comparison stimulus yielding the lowest value for  $D^2$ .

The computer code and the associated parameter files for running the model are available from the authors.

## II. EVALUATION OF THE MODEL

### A. Pitch of alternating phase harmonics

Carlyon and Shackleton (1994) presented listeners with harmonic complexes whose component harmonics were either in sine phase or alternating phase (sine/cosine/sine/cosine..., etc.). An alternating-phase stimulus is sometimes reported as having a pitch an octave above that of its corresponding sine-phase stimulus. It can be seen in Fig. 3 that the stimulus envelope changes when the phases are alternated; the period of the envelope is only half as long. If the pitch of an alternating-phase stimulus is reported to be double that of the corresponding sine-phase stimulus, the observation could be used as evidence in favor of models of pitch extraction which depend upon the period of the envelope (temporal theories). Spectral theories, on the other hand, cannot easily explain the rise in pitch because the spectral components of the stimulus have not changed.

Carlyon and Shackleton (1994) reported that the pitch appears to double for alternating-phase stimuli when they are restricted to a small number of *high* harmonics. On the other hand, if the stimuli are restricted to a small number of *low* harmonics, the effect disappears and both sine-phase and alternating-phase stimuli are reported as having the same pitch. The results appear to imply that two different pitch mechanisms are at work; one (phase insensitive) for low harmonics and the other (phase sensitive) for high harmonics. High-harmonic stimuli generate percepts consistent with temporal theories while low-harmonic stimuli generate percepts consistent with spectral theories.

By varying the fundamental frequency and the range of harmonics used in a stimulus, Carlyon and Shackleton (1994) were able to show that the critical variable was probably not the frequency of the harmonics but their “resolvability.” When harmonic components are well spaced relative to the width of auditory filters (Patterson and Moore, 1986), there is relatively little interaction of adjacent harmonics within a single filter and the components are said to be “resolved.” This result might be used to support a more specific argument that percepts based on *resolved* harmonics may be best accounted for using spectral theories, while percepts based on *unresolved* harmonics may be best accounted for using temporal theories.

The following demonstration shows that our original model can simulate the effect without further modification. To do this we used stimuli similar to those of Carlyon and Shackleton by creating alternating-phase stimuli using only

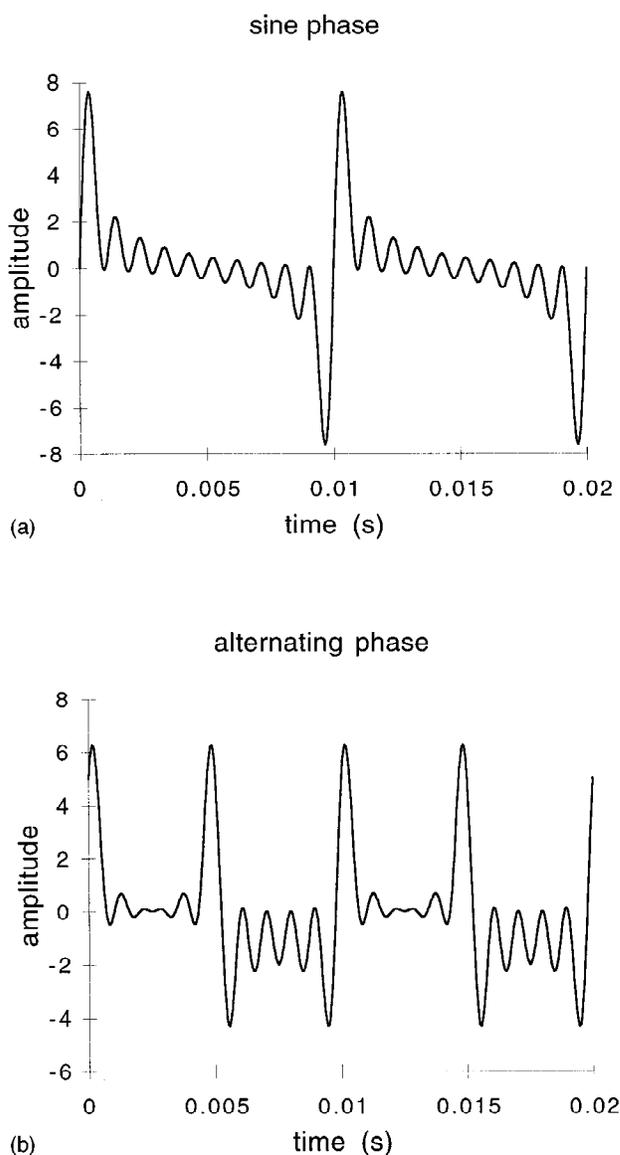


FIG. 3. Comparison of stimuli with harmonics 1–10 in (a) sine phase and (b) alternating phase.  $F_0 = 100$  Hz.

harmonics in limited-width frequency bands. The SACF for this alternating-phase stimulus was then compared with the SACFs for two similar stimuli whose harmonics were in sine phase; one with the same  $F_0$  and one with  $F_0$  set an octave higher.

The test stimulus was composed of harmonics in *alternating phase* with an  $F_0$  of 150 Hz. The two comparison stimuli were composed of harmonics in *sine phase* with  $F_0$  150 Hz and  $F_0$  300 Hz, respectively. The test was repeated three times using:

- (1) only LOW harmonics (range 125–625 Hz),
- (2) only MID harmonics (range 1375–1875 Hz), and
- (3) only HIGH harmonics (range 3900–5400 Hz).

The stimuli were generated using a signal consisting of 80 harmonics of the fundamental, digitally filtered using a Butterworth filter set to give skirts of 24 dB per octave. The range was used to specify the 3 dB down points of the filter.

The task is to find the best match between the test stimu-

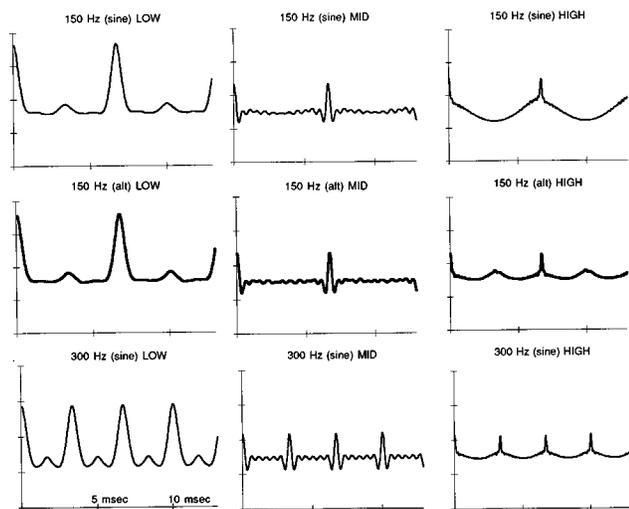


FIG. 4. The effect of alternating phase on the SACF. Top row; SACFs obtained using harmonics of 150 Hz in either LOW, MED, or HIGH ranges (see text). All harmonics are in sine phase. Middle row; as top row but the phases of successive harmonics alternate in phase. Bottom row; as top row but  $F_0 = 300$  Hz. Compare the SACF in the middle row (alternating phase) with the SACFs immediately above and below (sine phase). For LOW and MID, the best match is with the top row ( $F_0 = 150$  Hz). For HIGH, the best match is with the bottom row ( $F_0 = 300$  Hz).

lus (alternating phase) and either the 150-Hz or the 300-Hz comparison stimulus (sine phase). The SACFs for all nine stimuli are shown in Fig. 4, where the comparison can be made visually.

When LOW harmonics are used, the SACF of the alternating-phase 150-Hz test stimulus is indistinguishable from that of the 150-Hz sine-phase comparison stimulus. When HIGH harmonics are used, the SACF of the alternating-phase 150-Hz test stimulus generates a distinct extra peak that is missing from the sine-phase version. This makes it more similar to the SACF of the 300-Hz sine-phase comparison stimulus (bottom row). The model, therefore, predicts that a 150-Hz alternating-phase stimulus will be matched with 300-Hz sine-phase comparison stimulus when HIGH harmonics are used but not when LOW harmonics are used. This is essentially the same result as that reported by Carlyon and Shackleton (1994).

An explanation of the model's behavior can be inferred from the differences among the ACFs in Fig. 2. Figure 2(a) and (b) shows the ACFs and SACF for LOW and HIGH harmonics in *sine phase* with a fundamental frequency of 100 Hz. Contrast these figures with those given in Fig. 2(c) and (d) which is based on stimuli that are identical in all respects except that the components are in *alternating phase*. For LOW harmonics, the ACFs are indistinguishable between sine and alternating phase [2(a) and (c)]. This is because the harmonics are resolved by the filters so that the input to the individual ACFs are approximately pure tones. The ACF of a pure tone is not affected by the phase of the tone. As a consequence, changing the phase of components that are resolved by the filter bank has no effect on the pattern of ACFs and, by implication, has no effect on the percept. In the case of HIGH harmonics, on the other hand, the ACFs are different in the two corresponding figures [2(b)

and (d)]. For example, the pronounced dip at a lag of 3.3 ms for sine-phase stimuli is replaced by a distinct peak for cosine-phase stimuli. This occurs because the filters no longer resolve individual components. The within-channel waveform for multi-component signals is, therefore, strongly affected by the phase relationships of the interacting components. Autocorrelation is strongly affected by changes in shape of the waveform and this is reflected in the ACFs which predict that relative phase will be perceptually relevant when the components are unresolved (see also Patterson, 1987).

We may conclude from this demonstration that the model's behavior is consistent with the reports of listeners. Furthermore, it is not necessary to postulate two separate mechanisms to explain different percepts for resolved and unresolved harmonics when presented in alternating phase. It is a property of the four-stage model that it will respond differently to these two types of stimulus without further amendment.

The differences between the SACFs for the two stimuli are manifest along the length of the SACF and cannot be characterized simply in terms of a shift in the main peak in the SACF. When modeling the results of a number of psychophysical pitch studies, the location of the main peak in the SACF was used with some success as a predictor of pitch judgments (Meddis and Hewitt, 1991a). When modeling sensitivity to differences in phase relationships among harmonic components, comparisons along the length of the SACFs of the two stimuli were used (Meddis and Hewitt, 1991b). The latter procedure is a more general version of the former and probably a more secure approach. In the remaining evaluations of the model, comparisons will be made along substantial segments of the SACF.

## B. Pitch discrimination

Houtsma and Smurzynski (1990) showed that the frequency difference limen for the pitch of an 11 component harmonic complex increased with the number of the lowest harmonic in the complex. Carlyon and Shackleton (1994) found that this is a function of the *resolvability* of constituent harmonics. We explored this issue by testing the model using Carlyon and Shackleton's (1994) paradigm. We used stimuli with harmonic complexes containing either resolved (LOW, from 125 to 625 Hz) or unresolved (HIGH, from 3900 to 5400 Hz) harmonics. Comparisons were made between tones with  $F_0$ 's of 100 and 102 Hz. If the model is to simulate Carlyon and Shackleton's result we expect the 2% shift to produce a bigger change in the SACF when using low harmonics than when using high harmonics. To test this we calculated  $D^2$  between the SACFs for the 100-Hz and the 102-Hz tones. The stimuli were generated using a signal consisting of 80 harmonics of the fundamental digitally filtered using a Butterworth filter set to give skirts of 24 dB per octave. The range was used to specify the 3 dB down points of the filter. The SACF was calculated after 10 complete cycles of the stimulus.

The SACFs for the LOW and HIGH harmonic stimuli are given in Fig. 5(a) and (c). The SACFs are very different in shape for the resolved (LOW) and unresolved (HIGH)

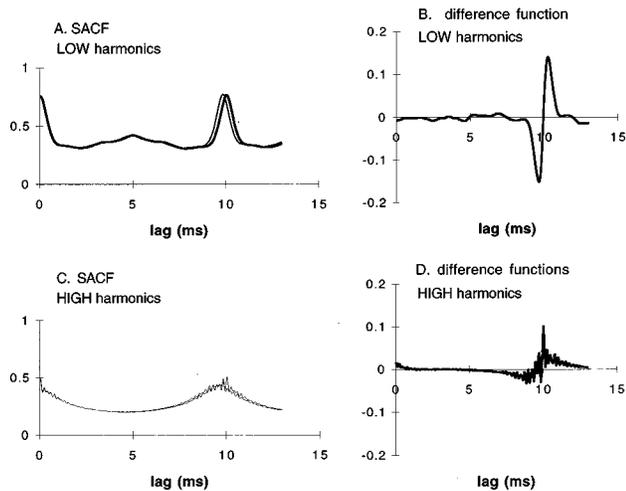


FIG. 5. SACFs and difference functions for signals with  $F_0$  of 100 and 102 Hz in the LOW and HIGH harmonic ranges. Top row, LOW harmonics; bottom row, HIGH harmonics. Left column, SACFs; right column, difference functions.

harmonics. The LOW stimulus has a narrow peak at a lag of 10 ms corresponding to a 100-Hz pitch percept. The HIGH stimulus has a broad and shallow peak at the same value.

A 2% shift in  $F_0$  creates very different effects for the two stimuli. For the LOW stimulus, a clear difference can be seen between the 100- and 102-Hz SACFs. For the HIGH stimulus, the difference is much less clear. Figure 5(b) and (d) shows the corresponding difference functions. The LOW stimulus produces a larger difference function when the pitch is shifted by 2%. Using the SACF range 80–1000 Hz,  $D^2$  was more than three times as great for LOW (0.37) than for HIGH (0.11) harmonics confirming our expectations.  $D^2$  was calculated here using a wide segment of the SACF (80–1000 Hz).

To demonstrate the generality and lawfulness of the effect, the above calculations were extended to include additional examples using a fundamental frequency of 200 Hz and two additional intermediate ranges of harmonics were studied. The four ranges are as follows: LOW (125–625 Hz), MID (1375–1875 Hz), MID/HIGH (2500–3500 Hz), and HIGH (3900–5400 Hz). Figure 6(a) shows the Euclidean distance for a 2% shift in  $F_0$  for all four harmonic ranges using two fundamental frequencies (100 and 250 Hz).

A large value of  $D^2$  implies a high discriminability of the two signals and therefore our  $D^2$  measure can be said to be loosely analogous to the  $d'$  statistic used in signal detection theory and employed by Carlyon and Shackleton. Our calculations based on the model are therefore qualitatively similar to their results (1994, Fig. 6) which show a deterioration in discriminability from LOW to HIGH ranges of harmonics.

Houtsma and Smurzynski (1990) measured frequency-difference limens using 11-component harmonic stimuli. They found an increase in threshold as the number of the lowest harmonic was raised from 7 to 19 for a 200-Hz fundamental. Their paradigm was simulated using an 80 harmonic stimulus band-pass filtered as described above; the lower and upper 3 dB down point of the filter were set to the

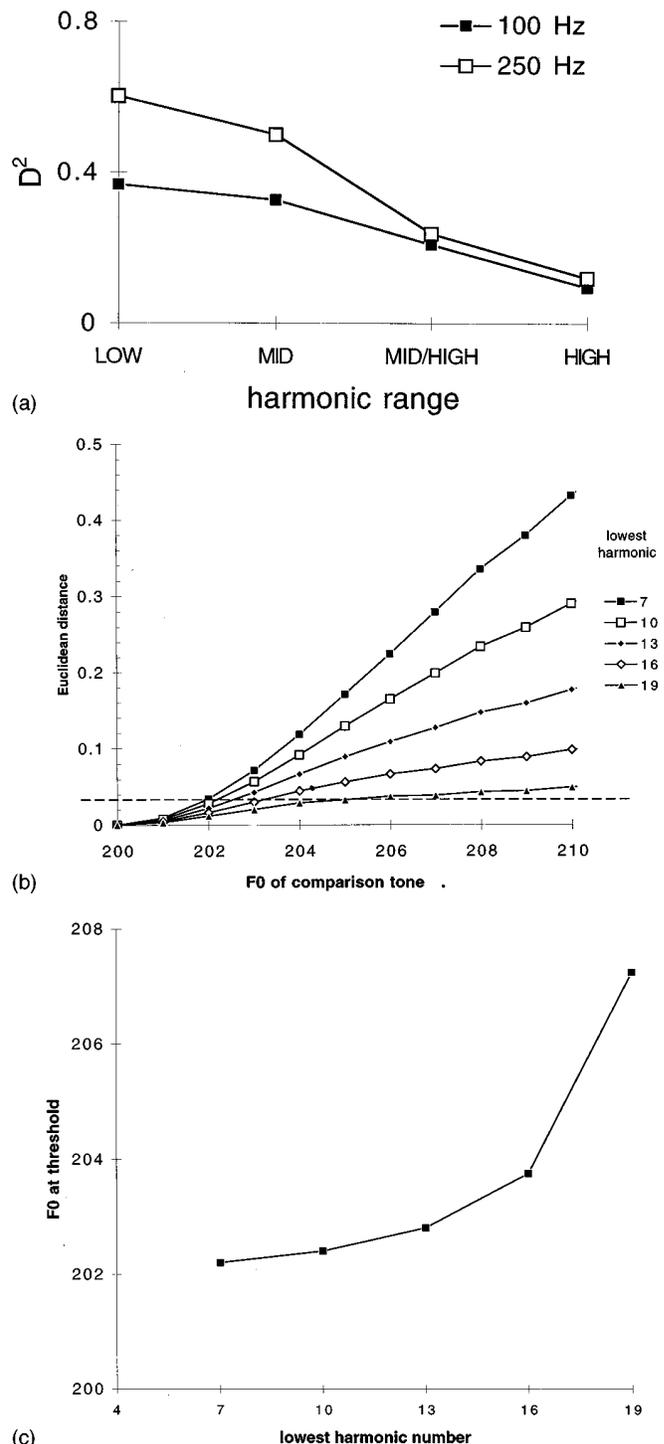


FIG. 6. (a)  $D^2$  for pairs of stimuli with fundamental frequencies differing by 2%. Four different ranges of harmonics and two values of  $F_0$  are used (see text for details). (b)  $D^2$  as a function of  $\Delta F_0$  and the number of the lowest harmonic. Tones consisted of 11 harmonics. The dashed horizontal line represents the criterion used to define "threshold." (c) "Threshold" values of  $F_0$  as a function of the number of the lowest harmonic. Values are derived from B using a criterion value of 0.042.

$n$ th and the  $(n + 11)$ th harmonic. Figure 6(b) shows  $D^2$  as a function of  $\Delta F_0$  for different ranges of harmonics. Stimuli with low harmonic number show steeper slopes than those with high harmonic number. The Euclidean distance calcu-

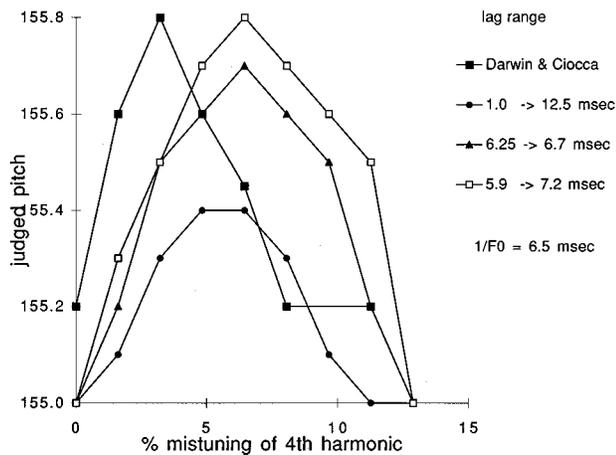


FIG. 7. Summary of model ‘best match’ decisions compared with Darwin and Ciocca’s (1994) results (open circles). The model results are repeated with different lag ranges.

lations were based on a lag range corresponding to  $F_0 \pm 10\%$ . When repeated using the wider range (80–1000 Hz), the pattern was similar.

The data in Fig. 6(b) can be represented more conventionally by setting a criterion value for  $D^2$ . A value of  $D^2 = 0.042$  was chosen to give thresholds similar to listeners and the model thresholds are shown in Fig. 6(c). These show the expected upward trend but do not replicate the leveling out observed by Houtsma and Smurzynski for the highest range of harmonics. The four-stage model generates shallower, less distinctive SACFs for stimuli with unresolved components. As a result small changes in  $F_0$  have less effect on the SACF than when resolved harmonics are used.

### C. The pitch of a complex tone with a mistuned harmonic

If one component of a harmonic complex is mistuned slightly, it gives rise to a shift in the pitch of the whole complex (e.g., Moore *et al.*, 1985). This effect is typically maximal for a 3%–4% mistuning of a low-numbered harmonic. The pitch shift can be as great as 1%. Darwin *et al.* (1994) have recently presented some highly detailed results for a harmonic stimulus with a fundamental frequency of 155 Hz. The behavior of the model is first illustrated using their paradigm.

To simulate Darwin *et al.*’s procedure, the SACFs were calculated for 10-component harmonic test tones of 100-ms duration with an  $F_0$  of 155 Hz and the fourth harmonic mistuned to a varying degree. To establish the best pitch match we also calculated the SACFs of a range of comparison 10-component harmonic tones, with  $F_0$  ranging between 155 and 156 Hz in steps of 0.1 Hz. For each test stimulus with a mistuned harmonic, the Euclidean distance was computed between its SACF and the SACFs of all comparison stimuli across the lag range corresponding to  $F_0 \pm 10\%$  (140–170 Hz). The comparison stimulus yielding the lowest value of  $D^2$  was judged to be the best match.

Figure 7 summarizes these results by showing the frequency of the best pitch match as a function of the degree of mistuning of the fourth harmonic and compares it with data

from Darwin *et al.* (1994). Figure 7 also shows the results of repeating the same process using different lag ranges along the SACF. Changing the lag range causes quantitative changes but leaves the qualitative results intact. The biggest pitch shifts were obtained using a lag range of 5.9–7.1 ms ( $1/F_0 \pm 10\%$ ). This range will be used for the rest of this section. Darwin *et al.* obtained pitch matches of 155.2 Hz for no mistuning (when the fourth harmonic is 620 Hz). This represents a small bias of 0.2 Hz in the listeners’ results. Allowing for this bias, the model results and the real data are in close agreement for the first three points.

When mistuning increases beyond 3%, the listeners’ pitch matches decline in frequency while the model’s pitch matches continue to rise in frequency. This may be attributable to an additional attentional mechanism beyond the scope of this study. As mistuning increases, listeners normally report that the mistuned harmonic can be heard as a separate entity. It is likely that this process of ‘hearing out’ a mistuned component is accompanied by a weakening of the contribution of that component to the pitch percept (Darwin, 1992). We shall not pursue this mechanism further beyond noting that we have experimented elsewhere with a model for segregating sounds using pitch to guide the process (Meddis and Hewitt, 1992).

The return of the *model*’s best match predictions to baseline levels when the mistuned harmonic is set to 700 Hz is a consequence of the fact that a mistuning of 700 Hz is both an upward mistuning of a fourth harmonic or an equal downward mistuning of a fifth harmonic; leading to a canceling of the effect.

### D. The dominance of individual harmonics

Moore *et al.* (1985) explored the magnitude of the pitch shift following the mistuning of the first to the sixth harmonic. The results were variable across subjects but it was possible to conclude that mistuning the *second* harmonic by 3% had the greatest effect on the pitch of the complex for  $F_0$  of 100, 200, and 400 Hz [Fig. 8(b)]. The higher the harmonic (above the second) the less its influence. By contrast, the first harmonic was more influential at higher fundamental frequencies. We, therefore, extended the above evaluation to study the effect of mistuning the first to the sixth harmonics.

To avoid complications related to the ‘hearing out’ phenomena discussed above, the evaluation is confined to 3% mistunings. The resulting pitch shifts are given in Fig. 8(a). The results were obtained using signal durations that were a complete number of cycles of the fundamental and close to 100 ms.  $D^2$  calculations were based on lag ranges corresponding to  $1/F_0 \pm 10\%$ .

While the observed shifts are quantitatively lower than Moore *et al.*’s results, the key characteristics of Moore *et al.*’s data are reproduced by the model. Mistuning the first harmonic has very little influence on the pitch of the complex with an  $F_0$  of 100 Hz but is increasingly influential at higher fundamental frequencies. The second harmonic is very influential for all three fundamental frequencies. Finally, the higher the harmonic (above the second) the less its influence on pitch.

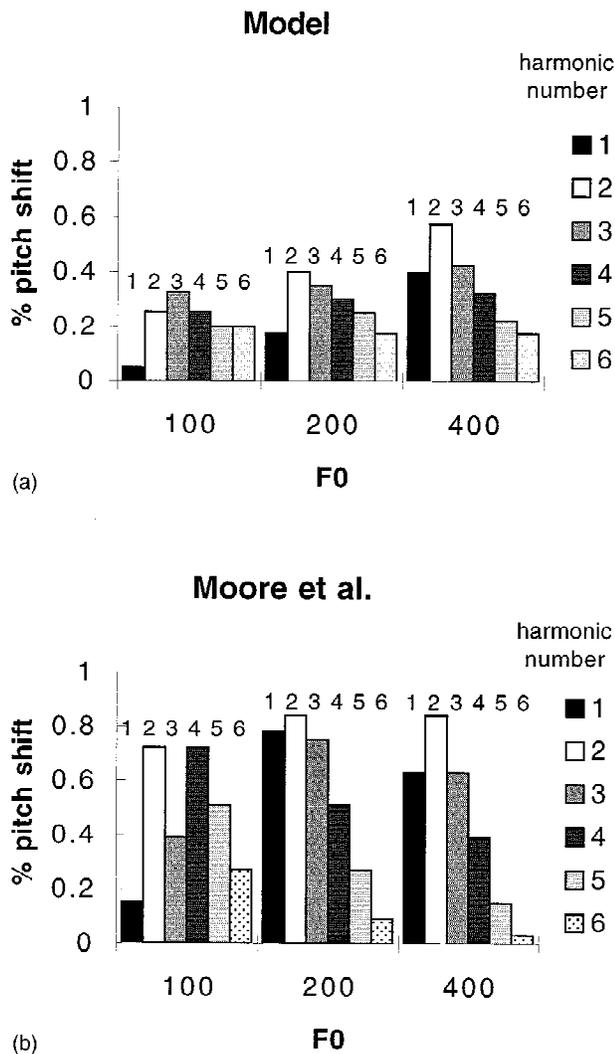


FIG. 8. Pitch shifts associated with 3% mistuning of harmonics 1–6 of 12-harmonic tones with  $F_0=100, 200,$  and  $400$  Hz. (a) Model results; (b) Moore *et al.*'s (1985) results.

It is difficult to explain the discrepancy in size between the results of Moore *et al.* and the model output. The model's results are consistent with those of Darwin *et al.* (see above) who also produce smaller pitch shifts than Moore *et al.* for corresponding stimuli.

The very small pitch shifts found for the model when the first harmonic was mistuned are not untypical of the empirical results for an  $F_0$  of 100 Hz. We speculated that this might arise from the action of the model's pre-emphasis function which attenuates low-frequency components. However, when this was switched off, the effect remained intact.

Moore *et al.* (1985) explained their experimental results in terms of the principle of a "dominance region." Ritsma's (1967, 1970) studies had earlier indicated that the third, fourth, and fifth harmonics made a dominant contribution to the pitch of a complex tone. Plomp (1967) suggested that the number of the dominant harmonic declined as the fundamental frequency increased. The model's behavior is consistent with Ritsma's dominance region except that we need to include the second harmonic in this region. Plomp's ideas are also represented in the model results; the third harmonic is

dominant when  $F_0$  is 100 Hz but the second harmonic is dominant at higher pitches. Similarly, the center of gravity of the three histograms moves toward lower harmonics as the pitch increases in both the model behavior and in the results of Moore *et al.*

The results in Fig. 8 were based on calculations using a lag range corresponding to frequencies within  $\pm 10\%$  of  $F_0$ . Additional calculations were made using a lag range corresponding to 80–1000 Hz. This resulted in some quantitative changes but the overall pattern of results and the conclusions that could be drawn were unchanged.

A careful examination of the model's operation leads to the suggestion that a number of principles interact to produce the dominance phenomenon. According to the model, the pitch sensation arises through the summation of the ACFs of the output of a bank of filters. At low frequencies each tone emerges from its corresponding filters as an unmodulated pure tone (at least for the fundamental frequencies studied here). The ACF of a pure tone can be represented as the cosine function. Low-frequency harmonics will have wider ACF peaks that will contribute more diffusely to the estimate of the most commonly occurring periodicity. According to this argument, higher-frequency harmonics should contribute more narrowly to the estimate and mistuning should disturb the pitch percept more. However, a second principle holds the first in check. As the frequency of the harmonic increases, the filtering which occurs at the level of the receptor potential reduces the ac component of its output. This reduces the amplitude of the cosine ACF and the extent of its contribution to the pitch percept. In Fig. 2 the peak-to-trough height of the ACFs can be seen to decrease in size with increasing filter center frequency. These two principles should produce an optimal harmonic number with reduced contributions from harmonics with higher and lower frequencies. This is certainly consistent with the results shown in Fig. 8. Because the channels are distributed on a logarithmic scale of frequency, high-frequency components are represented by a smaller number of channels and make a reduced contribution to the SACF. At the other end of the scale, the pre-emphasis of the outer and middle ear has the effect of reducing the amplitude of very low-frequency signals relative to middle-range frequencies above 1 kHz. While not critical to the principle of dominance, this latter effect will have some influence on the relative weights of stimulus components.

The principle of dominance, in the model at least, emerges from a set of no fewer than four subsidiary principles. It is doubtful whether the model is able to represent each effect accurately and, in any case, each principle will be subject to considerable individual differences in the real world.

### III. DISCUSSION

These results extend the range of successful predictions made by the four-stage periodicity model. In particular, these results are consistent with the view, strongly supported by recent research, that the pitch percept based on resolved harmonics will behave differently from that based on unresolved harmonics. A principle emerges that resolved har-

monics exercise a potent influence on the pitch percept but their relative phases do not influence the percept. By contrast, unresolved harmonics make a weaker contribution to the percept but their relative phases are perceptually salient. The ability of our model to simulate qualitative aspects of these phenomena suggest that it may not be necessary to posit two distinct mechanisms of pitch extraction.

The four-stage model represents a general approach to pitch extraction. Our particular implementation is only one of a number of alternatives. The use of linear gammatone filters to simulate peripheral mechanical filtering and the hair cell model to represent the half-wave rectification and low-pass filtering are both open to improvement. Similarly, the use of autocorrelation for the periodicity stage of the model may seem an unlikely choice if the process is considered to represent some physiological process. It is likely, however, that many alternative algorithms sensitive to periodicity could be substituted here without prejudice to the operation of the model. The final summation stage is also open to further study; the method used here is the simplest we could devise. Finally, our method of pitch matching by using the Euclidean distance between an arbitrarily long segment of the SACF is undoubtedly open to further development as the model is challenged by new data. At present, it remains unclear which segment of the SACF should be used for predicting pitch judgments. It is the basic four-stage nature of the model which is critical to our explanation of pitch rather than any specific implementation.

The ability of the system to detect periodicity appears to be limited to relatively long periods. This means that the system cannot extract information from the *fine structure* of the input waveform in the high-frequency channels. However, usable periodicity information is still available in amplitude modulation of the envelope of the output of high-frequency filters. It is, therefore, a matter of expediency that the signal be reduced to its envelope by half-wave rectification and low-pass filtering (stage ii). The low frequency characteristics of the envelope can then be processed by the periodicity detection system. A similar phenomenon can be observed in relation to our sensitivity to interaural time differences (ITDs) in continuous signals. At low signal frequencies these can be detected in the ITDs of the fine structure but at high frequencies this sensitivity is restricted to ITDs in the *envelope* of the signal (Henning, 1974).

In these two respects, the model can be seen to come close to making sense of Carlyon and Shackleton's (1994) proposal that there are two distinct pitch extraction mechanisms. High-frequency, unresolved harmonic components interact with each other and are stripped of fine structure information so that pitch judgments must be made on the basis of the signal envelope. Low-frequency, resolved harmonics do not interact and have no useful envelope information. The HIGH and LOW sets of components give rise to distinctly different SACF shapes (see Fig. 4); the former contribute to a shallow diffuse peak while the latter support narrower and higher peaks. This observation is consistent with their assertion that frequency-difference limens are exceptionally high when the comparison stimulus consists of high harmonics and the test stimulus consists of low harmonics; in terms of

SACFs, we are not comparing like with like. It is doubtful, however, that this amounts to two distinct mechanisms.

Pitch matching and difference threshold phenomena have been simulated in this study using comparisons along whole segments of the SACF. An alternative approach is to measure the location of the highest peak and look for differences in location between the two stimuli. In early (unpublished) tests we found this method to require an unreasonably high degree of resolution for the SACF. However, in Fig. 5(b) it can be clearly seen that the large differences between the two SACFs occur to the left and right of the main peak. The sensitivity of our method depends on including these parts of the SACF in our difference measure.

Despite its predictive power, autocorrelation is unlikely to be the method used by the nervous system to detect periodicity. Physiological studies have, however, shown that a number of different cell types in the cochlear nucleus act as band-pass AM amplifiers (Moller, 1976; Frisina *et al.*, 1990a, b; Kim *et al.*, 1990; Rhode, 1995; Hewitt and Meddis, 1994). These units can also phase lock to low-frequency pure tones. These two properties make it likely that they may be involved in the periodicity detection process. Some cells in the central nucleus of the inferior colliculus show an enhanced rate of firing to restricted ranges of frequencies of amplitude modulation (Langner, 1992; Rees and Palmer, 1989). It is possible, but not yet proven, that each cell of this type corresponds to a single point in our autocorrelation function. While it remains to be established that the correspondence is valid, it is important to stress that autocorrelation is merely substituting *pro tem* for the periodicity detector until the most appropriate physiological mechanism can be determined.

Our proposal falls short of a general theory of pitch. It does not, for example, address the issues of pitch variation with sound-pressure level, octave enlargement, and interaural pitch differences. Nor have we attempted to accommodate the phenomena of binaural pitch perception (Houtsma and Goldstein, 1972; Beerends and Houtsma, 1989; Cramer and Huggins, 1958). Similarly, it has not been possible to address Hartmann's (1994a, b) suggestion that inharmonicity is detected locally among adjacent harmonics and cannot be detected if adjacent harmonics are missing.

## ACKNOWLEDGMENTS

This research was supported by a grant from the Image Interpretation Initiative of the Science and Engineering Research Council, UK (GR/H52634). We are grateful to B. C. J. Moore, C. J. Darwin, E. Lopez-Poveda, and two anonymous reviewers for useful comments on an earlier version of this manuscript.

## APPENDIX: PARAMETERS USED IN THE EVALUATION OF THE MODEL

Signal parameters

- 1.00E-01 Stimulus signal duration (seconds)—variable.
- 5.00E-05 Stimulus sampling interval,  $dt$  (seconds).
- 2.500E-03 Ramp up rise time for signal (seconds).

Pre-emphasis filter

- 2 Filter order.

450.0 Lower, 3 dB down cutoff frequency (Hz).  
 8500.0 Upper, 3 dB down cutoff frequency (Hz).

#### Gammatone filter parameters

4 Order for the gamma tone filters.  
 100 Lowest center frequency of gamma-tone filter bank.  
 8000 Highest center frequency of BM filter bank.  
 60 No. of filters.

#### Hair cell parameters

100.0 Permeability constant  $A$  (units per second).  
 6000.0 Permeability constant  $B$  (units per second).  
 2000.0 Release rate  $g$  (units per second).  
 5.05 Replenishment rate  $\gamma$  (units per second).  
 2500.0 Loss rate  $l$  (units per second).  
 66.31 Reprocessing rate  $x$  (units per second).  
 6580.0 Recovery rate  $r$  (units per second).  
 1.0 Max. no. of transmitter packets in free pool  $m$ .  
 50 000.0 Firing rate scalar  $h$  (spikes per second)

#### Autocorrelation parameters

$13.0e-3$  maximum lag,  $l$  (seconds).  
 $10.0e-3$  Time constant,  $\tau$  (seconds).

<sup>1</sup>LUTEar is a computational library produced to simplify and standardise the implementation of auditory simulation investigations. It is written in ANSI C in a modular format. The library contains a large selection of auditory models and analysis routines. It is available on application to the authors or directly from this WWW site: <http://www.essex.ac.uk/psychology/hearinglab/>

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.

Berends, J. G., and Houtsma, A. J. M. (1989). "Pitch identification of simultaneous diotic and dichotic two tone complexes," *J. Acoust. Soc. Am.* **85**, 813–819.

Brown, G. J., and Cooke, M. (1994). "Perceptual grouping of musical sounds: A computational model," *J. New Music Res.* **23**, 1–25.

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

Cramer, E. M., and Huggins, W. H. (1958). "Creation of pitch through binaural interaction," *J. Acoust. Soc. Am.* **30**, 413–417.

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

Frisina, R. D., Smith, R. L., and Chamberlain, S. C. (1990a). "Encoding of amplitude modulation in the gerbil cochlear nucleus: I. A hierarchy of enhancement," *Hearing Res.* **44**, 99–122.

Frisina, R. D., Smith, R. L., and Chamberlain, S. C. (1990b). "Encoding of amplitude modulation in the Gerbil Cochlear Nucleus. II. Possible neural mechanisms," *Hearing Res.* **44**, 123–141.

Glasberg, B. R., and Moore, B. C. J. (1990). "Derivation of auditory filter shapes from notched-noise data," *Hearing Res.* **47**, 103–138.

Goldstein, J. L. (1973). "An optimum processor theory for the central formation of the pitch of complex tones," *J. Acoust. Soc. Am.* **54**, 1496–1517.

Hartmann, W. M. (1994a). "Hearing a mistuned harmonic in an otherwise periodic complex tone," *J. Acoust. Soc. Am.* **88**, 1712–1724.

Hartmann, W. M. (1994b). "On the perceptual segregation of steady-state tones," in *A Biological Framework for Speech Perception and Production*, edited by H. Kawahara (ATR Human Information Processing Laboratories, Kyoto, Japan).

Henning, G. B. (1974). "Detectability of interaural delay in high-frequency complex waveforms," *J. Acoust. Soc. Am.* **55**, 84–90.

Hewitt, M. J., and Meddis, R. (1994). "A computer model of amplitude

modulation sensitivity of single units in the inferior colliculus," *J. Acoust. Soc. Am.* **95**, 2145–2159.

Houtsma, A. J. M., and Goldstein, J. L. (1972). "The central origin of the pitch of complex tones: Evidence from musical interval recognition," *J. Acoust. Soc. Am.* **51**, 520–529.

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

Kim, D. O., Sirianni, J. G., and Chang, S. O. (1990). "Responses of DCN-PVCN neurons and auditory nerve fibers in unanesthetized decerebrate cats to AM and pure tones: Analysis with autocorrelation/power-spectrum," *Hearing Res.* **45**, 95–113.

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

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

Licklider, J. C. R. (1959). "Three auditory theories," in *Psychology: A Study of a Science*, edited by S. Koch (McGraw-Hill, New York).

Lyon, R. F. (1984). "Computational models of neural auditory processing," *IEEE Proc.* 1984. 36.1.1–36.1.4.

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 mechanical to 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). "Modelling the identification of concurrent vowels with different fundamental frequencies," *J. Acoust. Soc. Am.* **91**, 233–245.

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.

Moller, A. R. (1976). "Dynamic properties of primary auditory fibres compared with cells in the cochlear nucleus," *Acta Physiol. Scand.* **98**, 157–167.

Moore, B. C. J., Glasberg, B. R., and Peters, R. W. (1985). "Relative dominance of individual partials in determining the pitch of complex tones," *J. Acoust. Soc. Am.* **77**, 1853–1860.

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

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

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 in Hearing*, edited by B. C. J. Moore (Academic, London).

Patterson, R. D., Robinson, K. D., 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).

Plomp, R. (1967). "Pitch of complex tones," *J. Acoust. Soc. Am.* **41**, 1526–1533.

Rees, A., and Palmer, A. R. (1989). "Neuronal responses to amplitude modulated and pure-tone stimuli in the guinea pig inferior colliculus and their modification by broadband noise," *J. Acoust. Soc. Am.* **85**, 1978–1994.

Rhode, W. S. (1995). "Inter-spike intervals as a correlate of periodicity pitch in cat cochlear nucleus," *J. Acoust. Soc. Am.* **97**, 2414–2429.

Ritsma, R. J. (1967). "Frequencies dominant in the perception of the pitch of complex sounds," *J. Acoust. Soc. Am.* **42**, 191–198.

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

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

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).