Detection of across-frequency differences in fundamental frequency

Robert P. Carlyon
Laboratory of Experimental Psychology, Sussex University, Brighton BN1 9QG, Sussex, England

Laurent Demany1) and Catherine Semal
Laboratoire de Psychoacoustique, Université de Bordeaux II, 146 rue Léo-Saignat, F-33076 Bordeaux, France

(Received 1 November 1990; accepted for publication 27 August 1991)

Listeners discriminated between pairs of complex sounds, each consisting of two groups of components. Two harmonic complexes were played out through separate channels, and each filtered to obtain a “lower” and a “higher” group. The “carrier fundamental frequencies (F0s)” of both groups were usually 125 Hz; only those components in the lower group were resolvable by the peripheral auditory system. For the standard stimulus, the F0s of the two groups were frequency modulated coherently with each other, so that they were always equal. For the signal, the F0s of the two groups were modulated incoherently (π modulator delay), so that they differed by an amount that varied sinusoidally between values proportional to the depth of FM (the dependent variable). Stimuli were usually presented in continuous pink noise. The results showed that (i) when the components were added in sine or cosine phase, the mean threshold across listeners corresponded to a zero-peak modulation depth of 6%–7% (rms mistuning = 8.5%–10%); (ii) performance dropped to chance when the upper components were added in alternating sine-cosine phase, but was only moderately affected by the phase of the lower components; (iii) threshold for sine-phase stimuli improved by a factor of 1.6 when noise in the frequency region of the two component groups was removed; (iv) threshold increased moderately with increases in the frequency separation between the two component groups; (v) threshold dropped markedly when the F0s of both groups of components were increased so as to be resolvable by the peripheral auditory system; and (vi) performance dropped to chance when the nominal carrier F0s of the two groups of components differed from each other. It is concluded that listeners can perform simultaneous comparisons of F0s derived from resolved and unresolved harmonics, and that their performance on this task is fairly robust. Implications for the perceptual segregation of concurrent complex sounds, and for models of pitch perception, are discussed.

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

INTRODUCTION

There is a growing body of evidence that listeners can use differences in fundamental frequency (F0) between two simultaneous periodic sounds to perceptually separate those two sounds (Broadbent and Ladefoged, 1957; Scheffers, 1983; McAdams, 1984; U. T. Zwicker, 1984; Brokx and Nooteboom, 1982; Stubbs and Summerfield, 1988; Chalikia and Bregman, 1989; Assmann and Summerfield, 1989, 1990; Culling, 1990; Summerfield and Assman, 1991). For example, Scheffers (1983) reported that identification of simultaneous pairs of vowels improved with increasing differences between the F0s of the two vowels. F0 usually corresponds to pitch, and its role in the perceptual segregation of simultaneous sounds is one of the reasons for the continued interest in pitch perception, and has been the subject of several computational models (e.g., Parsons, 1976; Assmann and Summerfield, 1989, 1990; Meddis and Hewitt, 1991a,b; Slaney and Lyon, 1990).

The basis for many of these models comes from modern theories of pitch perception (e.g., Goldstein, 1973; Moore, 1989), which in turn are based mainly on experiments in which listeners are required to identify the pitch of a single sound source. Of particular relevance to the present study are experiments investigating the relative roles of two groups of harmonics, which differ in the degree to which they are resolved by the peripheral auditory system (Plomp, 1967; Ritsma, 1967; Moore et al., 1985). Such experiments are interesting because the two groups of components are assumed to convey F0 information in different ways: The lower, resolved, harmonics form the input to a central “pattern recognizer” which derives pitch from the relationship between the individual frequency components (Goldstein, 1973; Terhardt, 1974; Piszczalski and Galler, 1979; Terhardt et al., 1982), whereas the upper, unresolved, harmonics convey pitch information in the repetition rate of their composite waveform (Schouten, 1940, 1970). The data show that the low-frequency, resolved, harmonics are more important ("dominant") for pitch perception than are the high-frequency, unresolved, harmonics (Plomp, 1967; Ritsma, 1967; Moore et al., 1985). However, there is evi-
dence that listeners can also judge pitch from the purely tem-
poral information conveyed by unresolved components or by
amplitude modulated noise (Burns and Viemeister, 1981; 
Warren and Wrightson, 1981; Houtsma and Smurzynski, 
1990). Because of these findings, a number of authors have
proposed schemes involving a common mechanism for the
extraction of F0 information from both resolved and unre-
solved components. Such schemes include “autocorrelo-
gram” models (Licklider, 1951; Assmann and Summerfield, 
1990; Meddis and Hewitt, 1991a,b,c; Slaney and Lyon, 
1990), the qualitatively similar “crude sketch” outlined by
Moore (1989), and the “pulse-ribbon” model proposed by
Patterson (1989; Patterson et al., 1991). Recently,
Houtsma and Smurzynski (1990) have pointed out that the
model proposed by Srolovitz and Goldstein (1983) can also
derive pitch from both resolved and unresolved harmonics.1

Many of the newer models fall into the “autocorrela-
tion” category. They propose that the listener performs an
autocorrelation on the outputs of each of several overlapping
linear bandpass filters, with center frequencies (CFs) cover-
ing the audible range. Recent versions (Assmann and Sum-
merfield, 1990; Meddis and Hewitt, 1991a,b,c) state that the
individual autocorrelations are then summed to produce a
“summary autocorrelogram.” The channels passing the re-
solved components each produce a series of peaks in their
individual autocorrelograms at multiples of the period of
that component. When the autocorrelograms are summed
they produce a maximum at a period equal to l/F0. Chan-
nels that pass groups of unresolved components have peaks
in their individual autocorrelograms corresponding to mul-
tiples of the repetition rate of their outputs—also equal to
1/F0. Thus a multicomponent harmonic sound produces a
peak in the summary autocorrelogram at 1/F0 whether the
complex contains resolved harmonics, unresolved harmonic,
or both.

Regardless of the accuracy of specific models, the idea
that listeners can combine information from unresolved and
resolved harmonics is intuitively and ecologically appealing:
it provides a parsimonious explanation for a wide range of
pitch phenomena, and would supply a basis for listeners to
use a common F0 to group together components covering a
wide frequency range (Assmann and Summerfield, 1990;
Meddis and Hewitt, 1991a) state that the individual autocorrelations are then summed to produce a
“summary autocorrelogram.” The channels passing the re-
solved components each produce a series of peaks in their
individual autocorrelograms at multiples of the period of
that component. When the autocorrelograms are summed
they produce a maximum at a period equal to 1/F0. Chan-
nels that pass groups of unresolved components have peaks
in their individual autocorrelograms corresponding to mul-
tiples of the repetition rate of their outputs—also equal to
1/F0. Thus a multicomponent harmonic sound produces a
peak in the summary autocorrelogram at 1/F0 whether the
complex contains resolved harmonics, unresolved harmonics,
or both.

Darwin (1981) asked listeners to identify an ambiguous
four-formant sound, whose first three formants (F1,F2,F3)
formed the syllable /ru/, and of which F1, F3, and F4
formed the syllable /li/. Playing F2 on a different F0 to that
of the other formants led to an increase in “/li/” (F1,F3,F4)
responses, whereas playing F4 on a different F0 increased the
number of “/ru/” (F1,F2,F3) responses. Thus Darwin
showed that playing a formant on a different F0 to that of
the others made it less likely to contribute to the perception of
the composite sound than if all formants had the same F0.
He also showed, using a related task, that playing a formant
on an anomalous F0 led to an increase in the number of
perceived sound sources (see also Gardner et al., 1989).
However, his study was not concerned with the interaction
between resolved and unresolved harmonics, and it is possi-
bile that identification was based on the upper three for-
nants, which all contained components that were unre-
solved by the peripheral auditory system. It is also possible
that components of different formants interacted with each
other on the basilar membrane, so that identification was
influenced by within-channel, rather than across-channel
mechanisms.

Broadbent and Ladefoged (1957) overcame the prob-
lem of peripheral interactions by presenting listeners with
synthesized speech sounds made up of two formants, with
each formant played to a different ear. They reported that
even with this dichotic mode of presentation, listeners re-
ported hearing one “voice” when the two formants were
played on the same F0. When the formants were played on
different F0s, but to the same ear, listeners reported hearing
more than one “voice.” However, like Darwin, Broadbent
and Ladefoged were not interested in the extent to which the
two groups of components differed in their resolution by the
peripheral auditory system. The two formant frequencies
varied continuously, and it is likely that there were parts of
the speech signal during which both formants contained re-
solved components. They did perform a second experiment
with steady sounds consisting of pairs of steady-state for-
mants. However, as they presented their stimuli in quiet, and
used a speech synthesizer that produced formants with shal-
low slopes (Lawrence, 1953), it is possible that parts of
their upper formant contained resolved components, or that
parts of the lower formant contained unresolved compo-
nents.

The experiments reported here measured thresholds
(TΔF0s) for the detection of simultaneous differences in F0
between two groups of harmonics. The harmonics were cho-
osen so that one group would be well resolved, and the other
unresolved, by the peripheral auditory system. In most of
our experiments, the stimuli were filtered so that the har-
monic numbers of the most intense components of the lower
group were between 1 and 5, and those of the upper group
were between 11 and 15. Our choice of filter settings was
influenced by evidence that the resolvability of individual
harmonics decreases markedly as harmonic number is in-
creased from 5 to 11 (Plomp, 1964; Moore and Glasberg,
1983; Houtsma and Smurzynski, 1990); this difference in
resolvability was explicitly tested in experiment 1. Stimuli
were presented in a background of pink noise in order to
reduce the detectability of within-channel interactions
between the two groups of harmonics, and to limit the range of
harmonics in each group that were audible. The results show
that listeners can detect across-frequency differences in F0,
and that their performance is quite robust. TΔF0s were mea-
sured as a function of component phase, the difference in
frequency region occupied by the two groups of components,
signal-to-noise ratio (SNR), and baseline F0.

I. GENERAL METHOD
A. Rationale and preliminary experiment

In the initial procedure listeners were required to dis-
riminate between a steady harmonic complex tone (e.g.,
harmonics 1–5 and 11–19 of 125 Hz) and one in which the lower and higher groups of components were mistuned in opposite directions. This procedure was abandoned, but is mentioned here so as to clarify the rationale for the revised procedure, and because the new procedure and the way the data are plotted were influenced by the results obtained. The problems with the original procedure were that (i) mistuning was accompanied by changes in the overall spectral extent of the complex; (ii) the F0 had to be randomized from presentation to presentation, to prevent listeners from detecting mistuning by listening to the F0 of an individual group (even so, listeners might have erroneously latched on to the F0 of one group of harmonics); and (iii) with steady tones listeners do not have to perform a simultaneous comparison of the two groups of harmonics, even when the overall F0 is randomized: rather, they can simply switch their attention from one group to another and perform successive comparisons (Demany and Semal, 1990). Despite these problems, two findings were obtained that were reliable across listeners, and which are relevant to the present experiments. One of these was that $d'$ was roughly proportional to the percentage mistuning between the two groups of components; the other was that the adaptive threshold was about 10% for all listeners.

B. Main experiments: Signal generation and trial structure

The method of signal generation and the trial structure are shown in Figs. 1 and 2, respectively. In both the standard and signal intervals the first $N$ (usually 34) harmonics of a 125-Hz F0 were frequency modulated at a rate of 2.5 Hz and

![Diagram](image_url)
lus, and the extent of this variation was proportional to that of the modulation imposed on the two groups of components (Demany and Semal, 1988).

In both signal and standard intervals, the starting phase of the modulation was randomized from presentation to presentation. This was achieved by synthesizing an 800-ms sample of each stimulus and choosing a starting point for each 400-ms sample at random from the first 400 ms. The levels of all components with frequencies in the filter passbands were 45 dB SPL, and the complex tones were gated on and off with 5-ms raised-cosine ramps. A 5-kHz-wide pink noise was presented continuously. Its spectrum level in dB SPL was 17.8 at 500 Hz, 15.2 at 1000 Hz, 12.2 at 2000 Hz, and 8.8 at 4000 Hz.

All stimuli were presented through one earpiece of a Sennheiser HD414 headset, and were monitored using an HP3561A spectrum analyzer. To convert the voltage across the headphones at each frequency to the sound pressure levels quoted above, the sound pressure level in a B&K artificial ear (type 4153, B&K condenser microphone cartridge type 4134, external diameter 0.5 in.) in response to a 1-kHz sinusoid was measured. The spectrum was further shaped by the frequency response of the headphones, which, measured in the artificial ear, was flat between 125 Hz–1 kHz, and had gains of +3, +8, +6, 0, and –1 dB re: the output at 1 kHz at 2, 3, 4, 5, and 6 kHz, respectively. One advantage of using a noise background is that both signal and noise are shaped by the headphone response, so that the sensation level of the signal is not substantially affected by it.

The new paradigm is better than the original one described in Sec. I A for three reasons: (i) The spectral extent of signal and standard are always identical; (ii) there is no need to randomize the overall F0, as the range of F0s covered by each group of components is the same for standard and signal; and (iii) listeners are forced to make a simultaneous comparison of the continually changing F0s in each group.

C. Listeners and procedure

A total of five listeners took part in different experiments. Their absolute thresholds at octave frequencies between 250 and 8000 Hz were within 15 dB of the 1969 ANSI standard. Listener RC was the first author. Stimuli were presented using a 2I, 2AFC procedure with feedback. In most experiments thresholds were obtained using Levitt’s (1971) two-down one-up adaptive procedure, which converged on the 71%-correct ($d' = 0.78$) point on the psychometric function. The modulation depth was multiplied by 1.07 after every incorrect response and divided by 1.07 after every two consecutive correct responses, except for the trials before the first four turnpoints when a factor of 1.15 was used. Each run ended after 16 turnpoints and the threshold for each run was obtained from the geometric mean of the modulation depths at the last 12 turnpoints. Each threshold reported here is based on the geometric mean of six such runs. Geometric, rather than arithmetic, means were calculated because the preliminary experiments had indicated that $d'$ was roughly proportional to % mistuning, which in the present paradigm is proportional to modulation depth. It is conventional to average the logarithm of the variable that is proportional to $d'$ and to plot that variable on a log scale (e.g., one usually averages and plots signal energy in dB). Both of these conventions were followed.

In some experiments psychometric functions were measured, using the method of constant stimuli described by Carlyon (1991), and with each data point based on the average of 100 trials. Each of ten blocks consisted of ten trials at each modulation depth, preceded by six practice trials at the largest modulation depth used (20%). Listeners were tested individually in an IAC single-walled sound-attenuating booth within a large single-walled sound-attenuating room.

II. EXPERIMENT 1: EFFECT OF COMPONENT PHASE

A. Rationale

The purpose of experiment 1 was to provide an initial measure of the smallest detectable difference in F0 between a group of resolved components and a group of unresolved components, while controlling for a number of alternative cues. We were particularly concerned by two potential cues. First, despite the pink noise, listeners might detect beating between the (attenuated) components played out from the two DACs, for example those with harmonic numbers around seven and eight. Second, it is possible that the components in the upper group would be partially resolved by the peripheral auditory system. Below, we argue that both of these concerns can be addressed by independently manipulating the phase of the upper and lower components.

Figure 3(a) shows the output of a simulated auditory filter centered on 1675 Hz to the first 128 ms of an unmodulated complex, in quiet, with an F0 of 125 Hz and with components added in sine phase. The F0 can easily be identified from the regular repetition rate of the waveform. In Fig. 3(b) the input is the same, except that components are added in alternating sine-cosine phase. The output waveform has an ambiguous period and a much smaller peak factor, with the result that the F0 is hard to identify. In contrast,
phase does not markedly affect the output of filters with low CFs, which resolve the individual lower harmonics (Fig. 3(c) and (d); CF = 375 Hz). Given that listeners are less sensitive to across- than to within-channel phase differences (E. Zwicker, 1952; Patterson, 1987a, 1988), one would expect coding of F0 (and hence the detection of across-frequency differences in F0) to be affected less by the phase of the lower harmonics than by that of the upper harmonics. Consistent with this prediction, musical interval identification based on high harmonics deteriorates when the peak factor of the stimulus is reduced, whereas that based on low harmonics is unaffected (Houtsma and Smurzynski, 1990).

B. Method

Five conditions were used. In two conditions, termed "SINE" and "COS", all components were in sine or cosine phase, respectively. In a third condition ("ALT"), the odd-numbered components were in sine (0 deg) phase, and the even-numbered components were in cosine (90 deg) phase. In condition COS-ALT, harmonics 1-7 (played out from both DACs) were added in cosine phase, and harmonics 8-34 were added in alternating sine-cosine phase, starting with the eighth harmonic in sine phase. In condition ALT-COS harmonics 1-7 were in alternating phase and harmonics 8-34 were in cosine phase. Thresholds were measured for these five conditions using the adaptive procedure described in Sec. I C, except that in some conditions listeners could not reliably track a threshold smaller than a 30% initial modulation depth. Therefore 3-point psychometric functions (modulation depths = 5%, 10%, 20%) were measured in those conditions. The distinction between conditions in which a threshold could be tracked, and those in which it could not, was clear-cut: for the six trials of any condition, listeners could always track either greater than five, or fewer than two, thresholds.

In condition COS-ALT, psychometric functions were additionally measured with the pink noise replaced by a different noise, termed "barrier noise," which had energy attenuated in frequency regions corresponding to the passbands of filters 1 and 2. It was derived from two sources, which were summed. One was a 400-Hz-wide band of noise centered on 927 Hz and with a spectrum level of 16.2 dB SPL (equal to the level of the original pink noise at 927 Hz), generated by multiplying a 200-Hz low-pass noise by a 927-Hz sinusoid. The other source was the original pink noise high-pass filtered (48 dB/octave) at 2362 Hz, to mask the high-frequency slope of the excitation pattern. 3

C. Results

Thresholds estimated from the adaptive procedure are shown in Table I. Here and throughout, thresholds are expressed as percentage zero-to-peak modulation depths; the figures in parentheses are the values by which the thresholds should be multiplied and divided to obtain plus and minus one standard error. When comparing our results to those of studies that used static mistunings (e.g., Moore et al., 1986; Houtsma and Smurzynski, 1990), it is appropriate to consider measures such as the rms and maximum mistunings produced by our out-of-phase FM. These measures can be obtained by multiplying our thresholds by $\sqrt{2}$ and by 2, respectively.

Thresholds are quite similar for the sine and cosine stimuli, with reasonable agreement across listeners, and means of 7.0% and 6.4%, respectively. This is consistent with the fact that the outputs of auditory filters to the two types of stimulus are similar (checked using simulations similar to those in Fig. 3). For the ALT-COS stimulus some listeners' thresholds were elevated relative to those for the sine and cosine stimuli, but all listeners could perform the task and it was possible to track a threshold in each case. This was not true when the higher components were added in alternating phase (conditions ALT and COS-ALT), when it proved impossible to track a threshold. The reason for this is apparent in Fig. 4, which shows psychometric functions for conditions ALT, COS-ALT, and ALT-COS. Whereas the functions in condition ALT-COS (triangles) are monotonic and reach a $d'$ of at least 1.5 for each listener, those for conditions COS-ALT (squares) and ALT (circles) are fairly flat and usually hover close to chance ($d' = 0$). This is strong evidence that phase of the higher components is critical for the performance of the present task, but that the phase of the lower components is less so. This in turn suggests that listeners were detecting differences in F0 conveyed by resolved and unresolved components. Note that the magnitude of any beating between the two groups of components, such as between harmonics seven and eight, would have been similar in the COS-ALT and ALT-COS conditions.

An alternative explanation for the poor performance observed when the upper components were added in alternating phase is that the resulting low-peak factor at the output of high-frequency auditory filters might have reduced the detectability of the higher components: in contrast, a very peaky output may have been more detectable by "poking up" above the noise. This was the reason for measuring psychometric functions for condition COS-ALT in barrier noise. The inverted triangles in Fig. 4 show that even in this reduced noise, performance was much worse than for condition ALT-COS in pink noise.

III. EXPERIMENT 2: EFFECT OF FREQUENCY SEPARATION

A. Rationale

Many of the stimuli which listeners have to group on the basis of F0, such as the vowel sounds of speech, are broad-

<table>
<thead>
<tr>
<th>Phase</th>
<th>RC</th>
<th>JD</th>
<th>HC</th>
<th>JC</th>
<th>AB</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>SINE</td>
<td>6.0(1.2)</td>
<td>6.7(1.2)</td>
<td>5.0(1.1)</td>
<td>10.9(1.4)</td>
<td>7.0(1.1)</td>
<td>7.0</td>
</tr>
<tr>
<td>COS</td>
<td>5.6(1.1)</td>
<td>3.5(1.2)</td>
<td>5.8(1.2)</td>
<td>9.5(1.1)</td>
<td>9.5(1.3)</td>
<td>6.4</td>
</tr>
<tr>
<td>ALT-COS</td>
<td>8.4(1.1)</td>
<td>10.8(1.1)</td>
<td>5.2(1.2)</td>
<td>19.4(1.1)</td>
<td>24.4(1.2)</td>
<td>11.7</td>
</tr>
<tr>
<td>ALT</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>ALT</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
</tbody>
</table>

TABLE I. Thresholds obtained in the different phase conditions for each of the five listeners of experiment 1. Figures in parentheses are the numbers by which the thresholds should be multiplied and divided to obtain + / -- one standard error. "*" indicates that an adaptive threshold could not be tracked.
band. Experiment 2 investigated the frequency range across which listeners can compare F0s, and shifted the frequency region occupied by the upper components to progressively higher frequencies. In an attempt to check for any deterioration in the coding of F0 that might occur with increases in frequency region, frequency modulation thresholds (FMTs) for the upper group of components were measured at each filter setting.

B. Method

The stimuli were as described for the sine-phase stimuli in Sec. 1 B, except that the number of components in each group (before filtering) was increased to 50, and the settings of filter 2 were varied. The settings, chosen to be equal in bandwidth on a logarithmic scale, were 1375–1875 Hz, 1788–2438 Hz, 2325–3172 Hz, and 3025–4125 Hz. The pink noise bandwidth was increased to 10 kHz. FMTs were also measured for each of these filter settings. The outputs of DAC1 and filter 1 were disconnected from the headphone amplifier, and each trial consisted of a modulated (signal) and an unmodulated (standard) version of the higher harmonics. The starting phase of modulation was randomized from presentation to presentation, as it was for the TAFOs, and the step size and adaptive procedure were the same for the two types of measurement.

C. Results

TAFOs are shown for three listeners in Fig. 5. They show a moderate increase in threshold with increases in the cutoff of the upper filter, and hence with increases in the frequency separation between the two groups of components. The largest deterioration is for listener RC, whose threshold increases by a factor of 2.4 over the frequency region studied, which represents an increase in frequency separation, between the 3-dB down points of filters 1 and 2, from 750 to 2400 Hz.

FIG. 4. Psychometric functions for the different phase conditions of experiment 1. The horizontal dashed line corresponds to $d' = 0.78$, which is the value to which the adaptive procedures of experiments 1, 2, 3, and 4 converged.

FIG. 5. Thresholds (TAFOs) for experiment 2 as a function of the lower cutoff of filter 2. Error bars represent plus and minus one standard error.
The FMTs for the high-frequency group of components are shown in Fig. 6, and increase as the CF of that group is raised. This might suggest that part of the increase in $\Delta F0s$ was due to a degradation in the coding of $F0$ at high frequencies. However, in the next section we present evidence that FMTs for two groups of components do not provide a good prediction of the threshold for detecting a difference between their $F0$s. Hence, this conclusion may be unjustified.

The main result of experiment 2 is that $\Delta F0$s are fairly robust to increases in frequency separation between the two groups of components. This finding provides further evidence that listeners were not performing the task by monitoring the outputs of auditory filters, tuned between the two groups of components, that were responding to interactions between the two groups.

**IV. EXPERIMENT 3: EFFECTS OF SIGNAL-TO-NOISE RATIO**

A. Rationale

Experiment 1 provided an initial measure of the smallest detectable difference in $F0$ between two groups of resolved and unresolved components. In order to eliminate within-channel cues the stimuli were presented in a pink noise at a fairly low signal-to-noise ratio (SNR). It is known that the detection of FM in a single group of components is impaired by noise (Horst, 1989; Carlyon and Stubbs, 1989). Experiment 3 examined the extent to which the pink noise elevated $\Delta F0$s by degrading the coding of $F0$ changes in each frequency region. The general approach was to remove discrete frequency regions of the pink noise corresponding to the passbands of filters 1, 2, or both. Thresholds were then measured both for the detection of across-frequency $F0$ differences ($\Delta F0$s) and for the detection of frequency modulation of the lower ($FMT_L$s) and higher ($FMT_H$s) components.

![FIG. 6. Thresholds (FMTs) for experiment 2 as a function of the lower cutoff of filter 2. Error bars represent plus and minus 1 standard error.](image)

**B. Method**

For the main part of the experiment, the sine-phase stimuli of experiment 1 were used; where FMTs were measured the method was as described for experiment 2. Four types of noise were used, including the pink and barrier noises of experiment 1. In addition, a "high-frequency (HF)" noise was generated by high-pass filtering the pink noise at 927 Hz, and a "low-frequency (LF)" noise was generated by low-pass filtering the pink noise at 927 Hz and adding it to a high-pass-filtered (2362 Hz) version of itself. In all cases filtering was accomplished using two sections of a Kemper VBF/8 filter in series, with a combined attenuation rate of 48 dB/octave. $\Delta F0$s were measured using all four types of noise, $FMT_H$s were measured in the pink and LF noises, and $FMT_L$s were measured in the pink and HF noises. The $\Delta F0$s in pink noise were measured afresh for this experiment, to control for any effect of the practice listeners had had since experiment 1.

In an auxiliary experiment, FMTs were measured using the ALT-phase stimuli of experiment 1, in a pink noise background. In other respects, the method was the same as for the main experiment.

**C. Results**

The results of the main part of experiment 3 are shown in Table II. Comparison with the sine-phase data in Table I confirms that $\Delta F0$s in pink noise had not decreased significantly since experiment 1. For all five listeners $\Delta F0$s were lower for stimuli presented in barrier noise than for those presented in pink noise, indicating that the presence of noise in the frequency regions occupied by the two groups of components did indeed increase thresholds. However, the difference in thresholds between the two conditions was fairly small, corresponding to a mean factor of 1.6. It is also worth noting that FMTs for the low-frequency resolved group of harmonics were generally lower than those for the high-frequency unresolved group. (Compare the $FMT_L$s and $FMT_H$s in pink noise.) This is consistent with earlier data for the detection both of FM and of static frequency differences (Hoekstra and Ritsma, 1977; Hoekstra, 1979; Horst et al., 1984; Horst, 1989).

**TABLE II. FMTs and $\Delta F0$s for the different noise conditions of experiment 3. Standard errors are indicated in the same way as in Table I.**

<table>
<thead>
<tr>
<th>Noise Type</th>
<th>RC</th>
<th>JD</th>
<th>HC</th>
<th>JC</th>
<th>AB</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pink</td>
<td>5.2(1.1)</td>
<td>6.4(1.2)</td>
<td>6.2(1.1)</td>
<td>10.3(1.1)</td>
<td>7.4(1.1)</td>
<td>6.9</td>
</tr>
<tr>
<td>LF</td>
<td>4.6(1.2)</td>
<td>5.7(1.1)</td>
<td>2.3(1.2)</td>
<td>12.2(1.1)</td>
<td>4.7(1.2)</td>
<td>5.1</td>
</tr>
<tr>
<td>HF</td>
<td>6.4(1.1)</td>
<td>4.4(1.3)</td>
<td>4.0(1.2)</td>
<td>9.3(1.1)</td>
<td>5.2(1.1)</td>
<td>5.6</td>
</tr>
<tr>
<td>Barrier</td>
<td>3.0(1.1)</td>
<td>3.3(1.1)</td>
<td>3.1(1.2)</td>
<td>7.8(1.3)</td>
<td>6.2(1.1)</td>
<td>4.3</td>
</tr>
<tr>
<td>$FMT_L$</td>
<td>9.0(1.2)</td>
<td>7.5(1.1)</td>
<td>1.4(1.1)</td>
<td>4.8(1.1)</td>
<td>3.7(1.1)</td>
<td>4.4</td>
</tr>
<tr>
<td>LF</td>
<td>5.4(1.1)</td>
<td>6.3(1.2)</td>
<td>2.2(1.3)</td>
<td>2.5(1.1)</td>
<td>1.6(1.1)</td>
<td>3.1</td>
</tr>
<tr>
<td>$FMT_H$</td>
<td>1.2(1.1)</td>
<td>1.6(1.1)</td>
<td>2.3(1.2)</td>
<td>0.9(1.1)</td>
<td>1.4(1.1)</td>
<td>1.4</td>
</tr>
<tr>
<td>HF</td>
<td>0.9(1.1)</td>
<td>1.1(1.1)</td>
<td>1.4(1.1)</td>
<td>0.8(1.1)</td>
<td>1.1(1.1)</td>
<td>1.0</td>
</tr>
</tbody>
</table>
Table II shows that there was no clear relationship between the effects of removing a specific noise region on the performance in the FMT and TAFO tasks. For example, the TAFO for listener HC was lower in the presence of LF noise than in HF noise, indicating that removing noise in the high-frequency region improved performance more than did removing the low-frequency noise. Conversely, her FMT, was not reduced by the removal of high-frequency noise, whereas her FMT dropped when the low-frequency noise was filtered out. This shows that the relative influence of noise in the low- and high-frequency regions is not always the same for both tasks. Another relevant observation comes from the auxiliary experiment, which measured FMTs for the upper group of components with ALT-phase stimuli. For the two listeners who took part (RC and JC), FMTs for ALT-phase stimuli were similar to those for sine-phase stimuli, even though they could not detect across-frequency F0 differences when the upper components were in ALT phase. FMTs for the ALT-phase and sine-phase stimuli, with standard errors in parentheses, were 9.5% (1.05) and 9.0% (1.2) respectively for listener RC, and 5.0% (1.29) and 4.8% (1.1) for listener JC. Thus the phase of the upper components markedly affected thresholds in the TAFO task, but did not affect FMTs at all. It is also worth noting that listeners' percepts of the ALT-phase stimuli in the FMT task were qualitatively different from those for the sine-phase stimuli: for the ALT-phase stimuli listeners heard "something changing" during the signal (modulated) interval, but the percept was not of a modulation in pitch. For the sine-phase stimuli, the percept was of a clearly changing pitch. Both the data and listeners' introspections from experiment 3 indicate that there are cues available in the detection of F0 that are not useful in tasks, such as that used to measure TAFOs, that require the extraction of F0 information.

V. EXPERIMENT 4: EFFECT OF OVERALL F0

A. Rationale and method

Experiments 1–3 measured TAFOs for stimuli with an F0 of 125 Hz. The aim of experiment 4 was to examine how TAFOs varied with overall F0, with the passbands of filters 1 and 2 held constant. Previous research, for example on the identification of melodic intervals (Houtsma and Smurzynski, 1990), suggests that the encoding of the pitch of a complex sound is more accurate when its components are resolved than when they are unresolved. Accordingly, it is possible that at high F0s, when the upper as well as the lower group of components are resolved, TAFOs would decrease. However, this has not been demonstrated for simultaneous comparisons as required by the present task, and it is known that the discrimination of frequency relationships depends on whether the components are presented simultaneously or sequentially (Demany and Semal, 1990; Demany et al., 1991).

The method of stimulus generation and the procedure were the same as for the sine-phase condition of experiment 1, except that seven different F0s were used, and the number of components in each complex was adjusted for each F0 so that the highest carrier frequency was always at least 3250 Hz. This ensured that, even at a modulation depth of 20%, the highest component would always fall at least half an octave above the upper cutoff of filter 2 (27 dB down from passband). We did this because we wanted the width of the excitation pattern of the two groups of components to remain roughly constant throughout the modulation. The F0s used, in Hz, with the number of components in parentheses, were as follows: 62.5 (52), 77.5 (42), 100.0 (33), 125.0 (50), 157.5 (21), 197.5 (17), and 250.0 (13). In an additional condition, thresholds were measured with a 26-component 250-Hz complex and with the cutoffs of filters 1 and 2 doubled to 250–1250 and 2750–3750 Hz, respectively. Four listeners (RC, HC, JD, AB) took part in all conditions.

B. Results

TAFOs for four listeners are plotted as a function of F0 in Fig. 7. Thresholds decrease markedly at high F0s: the biggest decrease occurs as F0 increases from either 157.5 to 197.5 Hz (listeners RC, JD) or from 125 to 157.5 Hz (listeners HC and AB). The reduction in threshold between F0s of 125 and 250 Hz corresponds to a factor of 3.4 for listener RC, 6.1 for JD, 4.0 for HC, and 2.1 for AB. The pattern of results is consistent across listeners with minor exceptions. For three listeners, threshold is roughly constant for F0s at and below 125 Hz, but RC's thresholds decrease as F0 is reduced from 125 to 62.5 Hz. Listener AB's data show larger standard errors than those of the other listeners, and the smallest threshold reduction at high F0s.

The reduction in threshold at high F0s may be attributed to the upper group of components becoming resolved by the peripheral auditory system. F0 estimates based on resolvable components are more accurate than those based on unresolved components (Hoekstra and Ritsma, 1977;
per se. This assumption is supported by the data of Carlyon
1991). Unfortunately, the possibility remains that listeners
were primarily sensitive to FM incoherence, and could
detect it independently of any baseline mistuning, then
performance should be better in condition 100/225 than in
condition 100/100, because only in condition 100/225
would both groups of components be resolved. The assump-
tions underlying this prediction are discussed in Sec. VI C.

VI. EXPERIMENT 5: DETECTION OF F0 DIFFERENCES
OR FM INCOHERENCE?

A. Rationale

In experiments 1–4, we used coherently and incoherent-
ly modulated complex tones to investigate the detection of
across-frequency differences in F0. The assumption underly-
ing our paradigm is that listeners detect the F0 differences
caused by incoherent FM, rather than the FM incoherence
per se. This assumption is supported by the data of Carlyon
(1991), whose listeners could not detect across-frequency
FM incoherence in inharmonic sounds, and whose psycho-
metric functions for the detection of a simple mistuning im-
posed on one component of a harmonic sound could account
for the corresponding functions describing the detection of
FM incoherence. However, there are differences between the
stimuli used here and those used in the Carlyon (1991) ex-
periments, most notably that there are far more components
in each group in the present study. Without prolonging the
debate over whether FM incoherence can be detected under
any circumstances, it seems worth confirming that, in the
present study, listeners were indeed detecting instantaneous
differences in F0 rather than in FM coherence per se.

There are a number of findings from experiments 1–4
which suggest that our listeners were encoding the F0s of the
two component groups and comparing them in some way.
These include the dependence of thresholds on the phase of
the upper components reported in experiment 1, and the
reduction in thresholds at high F0s in experiment 4. This
latter finding was attributed to the greater resolution of the
components at high F0s than at lower F0s: as mentioned
earlier, comparisons of F0s are more accurate with resolved
than with unresolved harmonics (Houtsma and Smurzynski
1990). Unfortunately, the possibility remains that listeners
were detecting differences in the way that the F0s of the two
groups of components changed over time (FM inco-
herence), rather than instantaneous differences in F0 (mistun-
ing). In order to maintain an “FM incoherence” hypothesis,
one must assume that differences in performance across con-
ditions arise from differences in sensitivity to FM inco-
herence; thus one would have to assume from experiment 4 that
the detection of FM incoherence is better for resolved than
for unresolved harmonics. This prediction of the “FM inco-
herence” hypothesis was used in experiment 5 to distinguish
it from the assumption underlying experiments 1–4, that lis-
teners were detecting instantaneous differences in F0.

Two conditions were run. In condition 100/225 the low-
er group of components had a carrier F0 of 100 Hz, and the
higher group a carrier F0 of 225 Hz. In condition 100/100,
the carrier F0 of both groups was 100 Hz, and the stimuli
were similar to those of the 100-Hz F0 condition of exper-
iment 4. If listeners were sensitive to across-frequency differ-
cences in F0, then performance should be worse in condition
100/225 than in condition 100/100, because the two groups
of components are out of tune with each other on both the
standard and signal trials (cf. Carlyon, 1991). If, however,
listeners were primarily sensitive to FM incoherence, and
could detect it independently of any baseline mistuning, then
performance should be better in condition 100/225 than in
condition 100/100, because only in condition 100/225
would both groups of components be resolved. The assump-
tions underlying this prediction are discussed in Sec. VI C.

B. Method

The method of stimulus generation was similar to that
for the sine-phase condition of experiment 1, except as fol-
lows: in condition 100/225 the output from DAC 1 was a
frequency-modulated 33-component complex with an F0 of
100 Hz, and the output of DAC 2 was a frequency-modula-
ted 15-component complex with an F0 of 225 Hz. In the
standard interval, the outputs of the two DACs were modu-
lated coherently; in the signal interval the outputs were mod-
lated incoherently (τ modulator delay). Condition
100/100 was the same as condition 100/225 except that the
carrier F0 of the complexes played out through both DACs
was 100 Hz. Four-point psychometric functions were mea-
sured for each condition, with modulation depths of 2.5%,
5.0%, 10%, and 20%.

C. Results

Psychometric functions for listeners RC, HC, and JD
are shown in the three panels of Fig. 8. The results show that,
whereas in condition 100/100 the functions rise monoton-
ically with increases in modulation depth and reach a maxi-
mum of at least two for each listener, those in condition
100/225 are much flatter and indicate lower sensitivity. The
difference in performance between the two conditions is
larger for listeners RC and HC than for listener JD.

The results of experiment 5 confirm that listeners are
sensitive to at least the large across-frequency differences in
F0 used in condition 100/225; this F0 difference impairs per-
formance relative to condition 100/100. The data also show
that FM incoherence does not provide a sufficient basis for
detection, at least in the presence of a large F0 difference.
However, it could be argued that listeners were sensitive to
FM incoherence in experiments 1–4, but that the large mis-
tuning in condition 100/225 rendered this cue useless.
Whilst agreeing that experiment 5 does not provide suffi-
cient evidence to rule this out completely, we note that lis-
teners can detect differences in other “grouping” cues, even
phase at low frequencies. However, similar analyses of the outputs of filters with CFs between two adjacent and slightly higher harmonics \((e.g.,\) harmonics four and five) show beating between the harmonics at a rate equal to the separation between adjacent components, which, for these stimuli, is equal to \(F_0\). The beating is degraded, but not completely eliminated, by the pink noise. Thus listeners could have compared the rate of beating in the filters with low CFs to that in the filters responding to the higher group of harmonics. This possibility needs to be taken seriously because the shape of the waveform produced by interacting components is only markedly influenced by their phase relationship when there are more than two components interacting. Thus phase might have been relatively unimportant at low frequencies not because the components were resolved, but because listeners were attending to beating in filters that responded to only two components.

While stressing that we cannot rule out a role for beats in the present experiments, we have one reason for thinking that they are not essential for the detection of across-frequency differences in \(F_0\). Carlyon et al. \((1991)\) have shown that listeners can detect such differences when the \(F_0\) of the lower group of components is conveyed by only the odd harmonics of the fundamental. For such stimuli, beating in filters tuned between adjacent harmonics occurs at a rate of twice \(F_0\), so Carlyon et al.'s results indicate that it is not necessary for adjacent components to beat at a rate equal to \(F_0\) in order for listeners to detect across-frequency \(F_0\) differences. This conclusion is different from that of Bregman et al. \((1985)\), who presented listeners with complex tones consisting of two groups of three components. They reported that judgments of how "decomposed" the two groups sounded depended on the similarity of their AM rates, and not of their \(F_0s\) or pitches. In their study, the upper group always consisted of components at 1400, 1500, and 1600 Hz, and therefore had a pitch and an envelope repetition rate \((f_m)\) both equal to 100 Hz. They found that when this group was combined with one consisting of 392, 497, and 602 Hz \((f_m = 105 Hz, pitch \text{ of about } 100 \text{ Hz})\) the resulting complex sound was perceived as decomposed, but that, when it was combined with a group consisting of components at 428, 528, and 628 Hz \((f_m = 100 Hz, pitch \text{ of about } 105 \text{ Hz})\), a fused percept resulted.

An important difference between Bregman et al.'s study and that reported here is that their stimuli were presented at a higher level (approx. 64- to 74-dB/component) than ours (45 dB/component), and were presented in quiet whereas ours were presented in pink noise. It is possible that performance in their task was mediated by a stronger interaction \(\text{compared to our stimuli}\) between individual components of each group, or even by an interaction between the upper and lower components in channels tuned between them. Strickland et al. \((1989)\) reported that detection of across frequency envelope disparities at \(f_m = 128 \text{ Hz}\) was strongly affected by manipulations, such as presenting the two carriers to opposite ears or at very different levels, that restricted the use of within-channel cues. This was true even for frequency separations of the two carriers similar to that in the Bregman et al. study.

![Psychometric functions for conditions 100/100 (triangles) and 100/225 (squares) of experiment 5. The horizontal dashed line corresponds to \(d' = 0.78\), which is the value to which the adaptive procedures of experiments 1, 2, 3, and 4 converged.](https://example.com/psychometric_figs.png)
B. Effect of phase of lower components on threshold

Although the phase of the low-frequency components had a much smaller effect on performance than did that of the high-frequency components, it did have a significant effect for some listeners (Table 1). This difference, between thresholds in the sine- and alternating-phase conditions, was greatest for the two listeners with the highest thresholds (JC and AB), and requires explanation.

It is known that listeners can detect differences in the phases of components or of groups of components that drive different auditory filters, even though their sensitivity to such across-channel phase differences is less than that to within-channel differences (Patterson, 1987a, 1988). However, differences in across-channel phase are, presumably, perceived as differences in timbre, rather than in pitch, so it is not obvious how they would affect the present task. One possibility is that putting the lower components in ALT phase introduced an additional timbre difference between them and the upper components, thereby impairing listeners' ability to fuse the two groups. As the experimental task can be viewed as one of "grouping versus separation," this could have impaired performance slightly. Although we have no definite evidence for such an explanation, JC and AB's performance did improve when the adaptive procedure was changed to a constant-stimulus one: the thresholds estimated from their psychometric functions were 11.6% (listener JC) and 7.5% (AB), compared to adaptive thresholds of 19.4% and 24.4%. For the other three listeners, who were much less affected by the phase of the lower components, performance was less affected by the change in procedure. It is possible that, in the slightly easier paradigm, listeners JC and AB learnt to attend to the F0s of both groups of components, and to ignore irrelevant aspects of the stimulus such as timbre differences. Thus it may be that whereas the phase of the upper components affects pitch strength, and severely degrades performance for all listeners whatever the paradigm, that of the lower components affects timbre and only moderately degrades performance, and only for some listeners in some paradigms.

C. Implications for models of pitch perception

The data presented here indicate that listeners can detect differences between the F0s of two groups of components occupying different frequency regions, under conditions where only one group is resolved by the peripheral auditory system. The effects of varying the phase of the different component groups (experiment 1), and of changing the degree by which the upper components are resolved by the peripheral auditory system (experiment 4), are consistent with those of similar manipulations in pitch perception tasks (Plomp, 1964; Moore et al., 1986; Houtsma and Smurzynski, 1990). Our data are, to a first approximation, consistent with models of pitch perception that derive pitch estimates from both resolved and unresolved components. Such models include the "crude sketch" described by Moore (1989), "autocorrelogram" schemes (Licklider, 1951; Assmann and Summerfield, 1990; Meddis and Hewitt, 1991a), Patterson's (1987a,b, 1991) "pulse ribbon," and, as Houtsma and Smurzynski have pointed out, the model proposed by Srulovicz and Goldstein (1983). The use of frequency modulated stimuli means that, if listeners were comparing the outputs of two separate mechanisms—a pure "pattern recognizer" (e.g., Goldstein, 1973; Terhardt, 1974) and a purely temporal mechanism (Schouten, 1940, 1970), then the outputs of these two separate mechanisms would have to be rapidly converted to a form in which they could be compared. However, although it is parsimonious to conclude that listeners were analyzing the output of a single pitch mechanism, we note that the brain is capable of very fast conversions between different types of sensory information. These include the combination of interaural time and intensity cues in auditory lateralization (Stern and Colburn, 1978), and the combination of auditory and visual information in the perception of place of articulation (McGurk and MacDonald, 1976). Thus the strongest new conclusion to be drawn from our experiments is that, if there are two separate pitch mechanisms, their outputs must be rapidly converted to a mutually comparable form. Such a conclusion could not have been drawn from experiments using steady sounds, in which listeners could analyse the pitch of each complex in turn, and convert these measures to a common metric (Schouten, 1940; Plomp, 1967; Ritsma, 1967; Moore, 1977; Houtsma and Smurzynski, 1990). Note that such a strategy is possible even with simultaneous tones provided that they have a constant F0, in which case listeners can process each pitch in turn by switching their attention from one complex to the other (e.g., Broadbent and Ladevoged, 1957 [exp. 2]; see Demany and Semal, 1990).

Not all aspects of the present findings are consistent with the autocorrelogram models. It is known that, compared to high (unresolved) harmonics, low (resolved) harmonics produce stronger pitch perceptions, lower F0 discrimination thresholds, and are dominant for the perception of pitch (Plomp, 1967; Hoekstra and Ritsma, 1977; Hoekstra, 1979; Moore et al., 1985; Houtsma and Smurzynski, 1990). Meddis and Hewitt's (1991a) model accounts for the greater contribution to pitch perception of the low harmonics by the greater channel density at low frequencies. However, the results of experiment 4 indicate that TΔF0s drop markedly when F0 is increased so that both groups of components are resolved, even when the frequency region of the component groups (passbands of filters 1 and 2) are held constant. This indicates an improvement in the coding of F0 that cannot be attributed to an increase in channel density. If the greater dominance, pitch strength, and discriminability of resolved components arise from the same auditory process, then experiment 4 indicates that they are inconsistent with the explanation offered by Meddis and Hewitt's model.

D. Implications for perceptual sound segregation

Carlyon and Stubbs (1989) compared FMTs for harmonic sounds, inharmonic sounds, and "partially harmonic" sounds that each contained a subset of harmonically related components spanning a (different) restricted frequency region. They reported that, when the stimuli were presented in bursts of pink noise, FMTs were lower for the harmonic than for either the inharmonic or the partially harmonic sounds. They suggested that listeners can combine
The results presented here indicate that listeners can indeed extract such across-frequency information, which may be used to perceptually separate concurrent complex sounds.

The practical implications of the ability to compare F0s derived from resolved and unresolved components depend at least partially on the accuracy with which listeners can do this. In order to rule out peripheral interactions, our stimuli were presented in pink noise at a fairly low SNR. In real life the SNR is likely to be higher, at least over part of the frequency range (few noise sources mask as uniformly as does pink noise). Our best estimate therefore comes from the $T \Delta F_0$s measured in barrier noise, where the SNR was most favorable, and whose mean across listeners was 4.3%. This translates into an rms mistuning of 6.1%. These values are clearly larger than those for detecting mistuning of single resolved components (Moore et al., 1986; Demany et al., 1991), which are between 1% and 2% ($d' = 1$). However, this does not necessarily mean that across-frequency comparisons are not useful for the processing of concurrent complex sounds: thresholds for the detection of mistuning between components which are resolved from each other will be elevated by the presence of competing sound, whose low-frequency components may have frequencies very close to those of the target. Under such circumstances, across-frequency comparisons based on groups of components may aid segregation.

The size of our $T \Delta F_0$s are roughly consistent with the data of Culling (1990). He measured the identification of pairs of simultaneous vowels, and found that F0 differences between different formants of the same vowel impaired identification slightly, but only when the F0 difference was greater than one semitone (5.9%). Our data indicate that smaller F0 differences would not have been detectable.

E. Conclusions

The results presented previously show that listeners can detect across-frequency differences in F0. Performance is only moderately affected by increases in the frequency separation between the two groups of components, changes in the phase of the lower (resolved) components, and the presence of noise in the frequency region occupied by the two groups of components. The data are consistent with, but do not prove, models of pitch perception which propose that pitch is extracted from both resolved and unresolved components by a common mechanism. It is suggested that the combination of F0 information across a wide frequency region may be a useful strategy for the perceptual separation of concurrent harmonic sounds.

ACKNOWLEDGMENTS

This research was supported by a Royal Society University Research Fellowship and by a project grant from the Medical Research Council (U.K.) to the first author. The second author is affiliated with the Centre National de la Recherche Scientifique. We would like to thank E. Burns, A. de Cheveigné, J. Culling, C. Darwin, A. Q. Summerfield, and an anonymous reviewer for useful discussions and/or comments on previous versions of this article, and P. Russell for programming and signal-processing advice.

1 According to Surovich and Goldstein's (1983) model, the frequencies of individual harmonics are derived by passing the interspike interval histogram (ISIH) of each auditory nerve fiber through a filter matched to the CF of that fiber. In effect, the ISIH is multiplied by a squared cosine with peaks at integer multiples of 1/CF, and integrated. The output of the filters matched to each CF form the input to a "central spectrum," from which the F0 is estimated by a pattern recognition process (harmonic templates). Surovich and Goldstein (1983) concluded that "discrimination of periodicity pitch with closely-spaced harmonics...involves a completely different and much less efficient mechanism than frequency pattern recognition." (p. 1274). However, in pointing out that the model can in fact derive F0s from unresolved harmonics, Houtsma and Smurzynski (1990) note that, for fibers responding to unresolved harmonics, there will be a peak in the ISIH at a period corresponding to F0. This peak will be passed only by filters matched to harmonics of that F0 (which have periods equal to submultiples of 1/FO), and, accordingly, there will be peaks in the "central spectrum" at frequencies corresponding to unresolved (as well as resolved) harmonics.

2 The speech synthesizer used generated a formant of frequency F kHz by multiplying a periodic source by a 10-kHz sinusoid, filtering the resultant, and then multiplying it by another sinusoid of frequency (10 + F) kHz. This resulted in a formant with center frequency F and slopes equal (in dB/Hz) to the high-frequency rolloff of the original periodic source. Only a rough sketch of this source is given, resembling a repeated exponential that had decayed to about 11% of its initial (maximum) value by the end of each period. We synthesized the source using a period of 125 Hz, and calculated its power spectrum. This had a low-pass characteristic that dropped by 20 dB within 500 Hz, and by 26 dB within 1000 Hz, of its 0-Hz maximum.

3 The attenuation of filter 1 re: the level in its passbands, with the corresponding frequency in parentheses, is as follows: — 2.9 dB (750 Hz), — 3.4 dB (875 Hz), — 3.2 dB (1000 Hz), — 4.1 dB (1125 Hz), — 4.8 dB (1250 Hz). For filter 2, corresponding measures were — 42 dB (750 Hz), — 30.3 dB (875 Hz), — 21.2 dB (1000 Hz), — 13.3 dB (1125 Hz), — 6.7 dB (1250 Hz), — 5.2 dB (2000 Hz), — 8.6 dB (2125 Hz), — 12.3 dB (2250 Hz), — 16 dB (2375 Hz), — 19.5 dB (2500 Hz), — 22.9 dB (2625 Hz).

4 When manipulating the phase of the components of a complex sound it is necessary to take into account the phase response of the headphones, and also that of other analog equipment used, such as filters. Therefore the waveforms used in the simulations of Fig. 3 were obtained by playing the stimuli through the same equipment as used for the experiments, and measuring the output of the headphones with a B&K condenser microphone mounted in a B&K artificial ear (see Sec. I B for details). The microphone output was captured instantly using one of the ADC channels of the CED 1401 laboratory interface that also incorporated the DACs. The waveforms were then processed offline using an implementation of the gammatone auditory filter model described by Patterson et al. (1988). The resulting simulations were almost identical to similar ones applied directly to the calculated waveforms of the stimuli. The reason that the analog equipment did not markedly affect the phase manipulations used here is that the phase responses of most pieces of equipment change smoothly across frequency (except near sharp resonances), whereas the alternating phase used here resulted in abrupt phase changes between adjacent components.

5 Although the width of the excitation pattern of the stimuli remained approximately constant throughout the modulation, at the lowest F0s produced during the largest modulations, there were insufficient components to fill the upper skirts of filter 2. The high-frequency portion of the noise masked components in the highest frequency regions of these skirts.


