Discriminating between coherent and incoherent frequency modulation of complex tones

Robert P. Carlyon

Citation: The Journal of the Acoustical Society of America **89**, 329 (1991);
View online: https://doi.org/10.1121/1.400468
View Table of Contents: http://asa.scitation.org/toc/jas/89/1
Published by the Acoustical Society of America

Discriminating between coherent and incoherent frequency modulation of complex tones

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

(Received 12 February 1990; accepted for publication 22 August 1990)

A series of experiments measured the discrimination by human listeners of frequency-modulated complex tones which differed only in the coherence of frequency modulation (FM). For the coherently modulated tones all components were modulated by the same 5-Hz sinusoid, and by the same percentage of their starting frequencies, whereas for the incoherently modulated tones the modulation of one (target) component differed from that of the rest. When the 400-ms complex was composed of consecutive harmonics of a common fundamental, performance improved monotonically with increases in modulator delay, and was nearly perfect at the longest delays. When the complex was inharmonic, performance was near chance at all modulator delays, both for component frequencies between 1500 and 2500 Hz, and for component frequencies between 400 and 800 Hz. It is argued that listeners detected incoherence in harmonic complexes by detecting the resulting mistuning of the target component. This conclusion was supported by the finding that listeners were usually at least as good at detecting a fixed mistuning of the center component of a harmonic complex as they were at detecting a modulator phase delay imposed on it. A final experiment, with a stimulus duration of 1 s and slower modulation rates, showed that listeners could detect incoherence for some inharmonic complexes. However, detection was worse than for harmonic complexes and was, it is argued, based on weak harmonicity cues. The results of all experiments point to the absence of an across-frequency mechanism specific to the detection of FM incoherence.

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

INTRODUCTION

A. Motivation

The ability of normal listeners to perceptually separate concurrent complex sounds reveals a sophisticated form of auditory processing. In many situations, such as that of attending to one speaker against a background of other voices, the signal and noise have similar spectral and temporal characteristics, yet listeners can understand speech at very unfavorable signal-to-noise ratios (Brokx and Nooteboom, 1982). Perceptual separation of competing voices is possible even in the absence of the binaural cues that arise from differences in spatial location between signal and noise (Cherry, 1953; Brokx and Nooteboom, 1982). The experiments presented here pertain to the use of monaural cues in the perceptual separation of concurrent complex sounds.

Perhaps the most established monaural cue is that of differences in fundamental frequency ($F0$) between different sound sources. $F0$ differences have been shown to be useful for the separation of simultaneous voices (Brokx and Nooteboom, 1982; Scheffers, 1983; Stubbs and Summerfield, 1988; Summerfield and Assmann, 1990), and have been utilized by speech-separation signal-processing algorithms (Parsons, 1976; Stubbs and Summerfield, 1988). In a recent paper, Carlyon and Stubbs (1989) reported that thresholds for detecting frequency modulation (FM) of a complex sound in noise were lower for a harmonic than for an inharmonic sound, even when the inharmonic sound contained a subset of harmonically related components in a restricted frequency region. They concluded that listeners combine $F0$ information across different frequency regions when detecting FM in noise. The experiments reported here were motivated by some pilot work investigating the role of another cue, that of *FM coherence*, in a similar task. FM coherence refers to the fact that when the $F0$ of a complex sound is modulated, all the components change frequency in the same direction at the same time. It seemed plausible that listeners could use FM coherence to extract the common frequency modulation of the components of a complex sound from an interfering noise. It soon became apparent that it was first necessary to measure sensitivity to FM coherence, and accordingly this paper describes experiments on discrimination between incoherently and coherently modulated complex sounds. The detection of FM incoherence is relevant to "real-life" situations in that it is a prerequisite for the further use of incoherence in the perceptual separation of concurrent complex sounds.

B. Review

Although listeners could theoretically use FM coherence to group together the coherently modulated components of a complex sound, evidence that they actually do so is, at best, ambiguous. McAdams (1989) asked listeners to judge the prominence of a vowel presented simultaneously with two other vowels. Although frequency-modulated vowels were judged to be more prominent than unmodulated vowels, prominence ratings were unaffected by whether the target vowel's $F0$ was modulated by the same or by a different waveform as those of the other two. Gardner and Darwin (1986) found that modulating the 500-Hz component of a vowel independently of the other components had no effect on listeners' phonetic categorization of that vowel, whereas *mistuning* a component does affect vowel categorization (Darwin and Gardner, 1986). Schooneveldt and Moore (1988) found that the detection of a tone in a frequency-modulated band of noise was not improved by the addition of a coherently modulated "flanking" band of noise, although coherently modulating the *amplitude* of different masker components does aid signal detection (Hall *et al.*, 1984).

There is, however, some evidence that listeners can use FM coherence to perceptually separate concurrent complex sounds. McAdams (1984) presented listeners with pairs of successive complex sounds, with one sound having all components modulated by the same waveform, and the other having one component modulated independently of the rest. Listeners reliably identified the incoherently modulated complex as containing more sources. Grose and Hall (1990) used a paradigm similar to that of Schooneveldt and Moore (1988), but modulated their masker and flanker bands over a greater frequency range. They found that thresholds for a tone masked by a frequency-modulated band of noise were slightly lower in the presence of two flanking bands of noise that were coherently modulated with the masker, than when the flanking and masker bands were modulated incoherently. Recently, Chalikia and Bregman (1990) measured the identification of pairs of simultaneous vowels whose $F0$'s were modulated by linear glides. They found that identication was slightly superior when the $F0$'s of the two vowels crossed each other during the glides than when the $F0$ contours were parallel. However, in most conditions of a previous study (Chalikia and Bregman, 1989), they failed to find significantly better identification with crossing than with parallel glides.

One possible reason for the discrepancies between different studies is that there are a number of ways in which listeners might process FM coherence, and the availability of each cue may vary across the different stimuli and paradigms used. Some of the different strategies that listeners might use are described below, and the way in which they were controlled for in the present experiments described.

C. Potential mechanisms

One way to detect FM incoherence would be to analyze the outputs of the auditory filters responding to each frequency component and to extract the direction and rate of change of each component. Listeners would identify a complex as being "incoherent" when two or more of its frequency components changed frequency in different directions. If listeners could use such a strategy, it would imply a genuine across-frequency mechanism that was specific to the detection of FM coherence. The main aim of the present study was to investigate whether this strategy was available to listeners.

A second strategy, which the present study concentrated on controlling for, is possible when the components of a complex are harmonically related. Incoherent FM of one component causes it to move in and out of tune with the other components, so listeners could detect the mistuning, rather than the incoherence *per se*. Therefore, both harmonic and inharmonic stimuli were used here. If subjects were simply detecting inharmonicity, then discrimination should be worse when both signal (incoherent) and nonsignal (coherent) sounds are inharmonic, than when the coherent sound is harmonic.

A third strategy arises from the fact that, as each component of a modulated complex sound moves through the passbands of successive auditory filters, the output of each filter is subjected to amplitude modulation (AM). When two components undergo coherent FM, filters tuned (for example) just below their starting frequencies will undergo coherent AM, and when the FM is incoherent, so will be the AM. Wakefield and Edwards (1989) have shown that listeners could detect FM incoherence in such a way when large modulation depths were used. The detection of FM-induced AM was restricted in the present study by the use of fairly small (usually $+/-$ 2.5%) modulation depths.

A fourth way in which listeners could detect FM incoherence would be via a within-channel mechanism. When two components are moderately close together in frequency, incoherent FM would cause them to move into and out of the passband of a single auditory filter, and to beat together irregularly. This irregular beating might be detected by the listener, and used to identify the signal (Schooneveldt and Moore, 1987). It is difficult to generalize the usefulness of within-channel cues in a detection experiment to real-life situations, where components arising from both the same and different complex sounds interact within a filter irrespective of any FM. Therefore, the stimuli used in the present study had components that were well resolved by the peripheral auditory system. A continuous pink noise was presented throughout all measurements to prevent listeners attending to filters with center frequencies between two adjacent components, whose responses might otherwise be strongly affected by the combined waveform of the two components. Peripheral interactions were also reduced by the fairly low level (45 dB SPL/component) of the stimuli.

In summary, a series of experiments was performed using both harmonic and inharmonic complexes to investigate whether listeners can detect across-frequency FM coherence independently of the mistuning that it causes. The use of fairly small FM depths, the fairly low stimulus level, and the presence of a continuous pink noise reduced the chances of listeners using either FM-induced AM or within-channel cues to detect FM incoherence. Most of the stimuli consisted of either two or three components, and fell in the frequency range 1500–2500 Hz, where thresholds for detecting FM are a constant percentage of signal frequency (Carlyon and Stubbs, 1989).

I. EXPERIMENTS 1–3: DETECTION OF MODULATOR PHASE DELAY

This section describes three experiments which used similar procedures and methods of signal generation. All experiments measured either three- or six-point psychometric functions for the detection of FM incoherence. Table I shows the frequencies of the carrier components, the number of points measured, and the FM depth used in each experiment. Note that, although each experiment included complexes that are referred to as inharmonic, harmonicity cues

TABLE I. Carrier frequencies, number of points measured, and FM depth used in experiments 1–3. The target component of each complex is italicized.

Experiment	Carrier frequencies (Hz)	No. of points	FM depth(%)
1	1500,*2000*,2500	6	2.5
	1500,*2000*		
	2000,2500		
	1600,*2300*		
2	1500,*2000*,2500	3	2.5
	1500,*1950*,2500		
	1500,*1900*,2500		
	1500,*2050*,2500		
	1500,*2100*,2500		
3	400,*600*	3	5.0
	400,*700*		
	400,*800*		

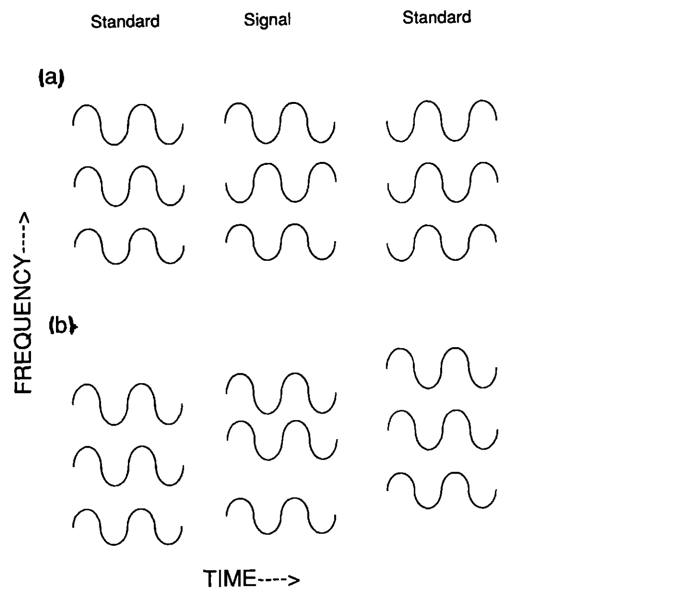

FIG. 1. (a) Schematic spectrogram of the trial structure in experiments 1–3, for a trial in which the signal occurred in the second interval (in 50% of trials the signal occurred in the third interval). (b) As above, for the mistuning condition of experiment 4. Note the different vertical location of each of the three complexes, reflecting the randomization of $F0$ from trial to trial. The trial structure for the incoherence condition of experiment 4 was similar to that shown in (a), except that $F0$ of the complex was randomized from trial to trial.

were not always completely eliminated. For example, when the 1600 + *2300* Hz complex of experiment 1 was modulated incoherently (π phase delay, 2.5% modulation depth), the two components moved in and out of a 2/3 frequency ratio. Although we will return to this point later, complexes will be described as "inharmonic" whenever the (unmodulated) carrier frequencies of their components do not form a simple harmonic ratio. The procedure and method of signal generation were common to all three experiments, and are described below.

A. General paradigm

Listeners were presented with triplets of 400-ms complex sounds in a 3I, 2AFC paradigm. A schematic representation of the trial structure is shown in Fig. 1(a). All stimuli were frequency modulated by two cycles of a 5-Hz sinusoid. In the first interval, the phase of the waveform modulating each component was identical, and this was also the case in either the second or third intervals (chosen at random). In the remaining (signal) interval, the phase of the waveform modulating one component (italicized in Table I) was delayed by amounts ranging from $\pi/6$ to π, and psychometric functions were measured relating d' to modulator phase delay.[1] The waveforms modulating the non-target components always started at a zero crossing, but the phase of all modulators was inverted on 50% of the trials in order to deter listeners from responding to the starting phase of the waveform modulating the target component (Carlyon and Stubbs, 1989).

B. Signal generation

The 400-ms signals were calculated, shaped with 5-ms raised-cosine ramps, and stored on disk. They were generated via a CED 1401 laboratory interface (12-bit amplitude quantization, 10 000-Hz sampling rate) under computer control, low-pass filtered (Kemo VBF 25.01, 135 dB/octave, 4300-Hz cutoff), attenuated (Wilsonics PATT), and mixed with a continuous pink noise via a headphone amplifier. The method of generation was similar to that reported by Carlyon and Stubbs (1989), and the formula describing the waveform was similar to theirs except for a term describing the modulator phase delay on one component:

$$y(t) = \sum_{n=1}^{N} A \sin\left[2\pi f_n t + \phi_n + g \frac{P}{100} \frac{f_n}{f_m} \cos(2\pi f_m + \phi_{nm})\right] \quad (1)$$

where N was the total number of carrier components, A was an amplitude constant, f_n was the center frequency of the nth component, ϕ_n was its starting phase, f_m was the frequency of the modulator (5 Hz in experiments 1–3), and ϕ_{nm} was the modulator phase for the nth component. In the nonsignal intervals ϕ_{nm} was zero for all components, and in the signal interval it ranged from $\pi/6$ to π for one (target) component and was zero for the others. The quantity g was + 1 on half of all stimulus presentations, and − 1 on the other half, corresponding to the modulation on the nontarget components starting at positive and negative zero crossings, respectively [Fig. 1(a)]. The quantity P was the maximum frequency deviation imposed on each carrier as a percentage of its starting frequency, and was 2.5 in experiments 1 and 2, and 5.0 in experiment 3. All stimuli were presented at a level of 45 dB SPL per component. The components were added in the low-peak-factor phase relationship described by Carlyon and Stubbs (1989).

A continuous pink noise was present throughout all measurements. It was generated by an analog noise generator with a built-in pink noise filter and a bandwidth set to 5 kHz, attenuated (Wilsonics PATT), and mixed with the signal. 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 monitored using an HP3561A signal analyzer and presented through one earpiece of a Sennheiser HD414 headset.

C. Listeners and procedure

Three listeners were tested individually in a sound-attenuating room. Listener RC was the author; the other two listeners had several weeks (2 h testing per day) of practice on the experimental task. Their absolute thresholds were within 15 dB of the 1969 ANSI standard at all audiometric frequencies. Psychometric functions were measured using the method of constant stimuli. In experiment 1, stimuli were presented in blocks of 120 trials [6 modulator delays $\times 10$ repetitions $\times 2$ modulator starting phases (positive/negative zero crossing)] in random order, preceeded by six practice trials at a π phase delay. Ten blocks of trials were run for each of the four conditions, so that for each combination of modulator delay and modulator starting phase there were 100 trials (ten blocks $\times$ ten repetitions per block). After a listener had completed ten blocks for a condition the total number of correct responses for each block was calculated, and a mean and 80% confidence limit across blocks obtained. Any block of trials falling outside the confidence limits was discarded, and a fresh block of trials obtained (Carlyon *et al.*, 1990). As it turned out, no significant differences were found between the data for the two modulator starting phases, so data were averaged across this variable (200 trials per datapoint). The procedure for experiments 2 and 3 was the same, except that each block consisted of only 60 trials (three modulator delays $\times$ ten repetitions $\times$ two modulator starting phases). LEDs provided feedback to the listeners after each trial.

D. Results

The results of experiment 1 are shown in Fig. 2. For the three-component harmonic complex (1500 + *2000* + 2500 Hz; inverted triangles) psychometric functions rose monotonically with increasing modulator phase delay, with a maximum value of d' between about 2 and about 3, depending on the listener. For this complex, then, all listeners obtained near-perfect performance at the longest modulator delay. The functions for the two-component inharmonic complex (1600 + *2300* Hz; upright triangles) were quite different: for all listeners and all modulator delays, d' did not exceed the value of 0.25 required for it to be significantly greater than zero.[2] Thus, for the inharmonic complex, listeners could not discriminate between two components that changed frequency together, and two that always changed frequency in opposite directions. Later experiments (e.g., experiment 3) will show that this breakdown in performance was not due to the wide frequency spacing of the components in the inharmonic complex. The breakdown occurred despite the fact that listeners could easily *detect* the 2.5% modulation used here: an auxiliary experiment using a paradigm similar to that of Carlyon and Stubbs showed that listeners could reliably detect a 0.8% FM imposed on a 1600- or 2300-Hz sinusoid.

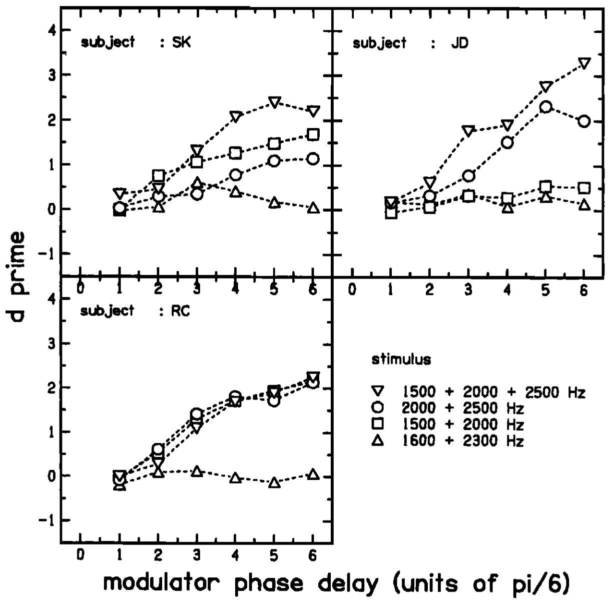

FIG. 2. Detectability (d') as a function of the modulator phase delay of the target component in experiment 1. Data are shown for one inharmonic (1600 + *2300* Hz) and three harmonic complexes. The abscissa is labeled in six intervals of $\pi/6$, up to a maximum of π.

For the harmonic two-component complexes, which had a rather weak pitch, performance varied across listeners: it was generally worse than for the three-component harmonic complex, but better than chance. The results suggest that, for the stimuli of experiment 1 at least, listeners can only detect FM incoherence by virtue of the mistuning from a harmonic relationship that it causes. When the carrier was inharmonic, performance was at chance.

The importance of harmonicity for the detection of FM incoherence is further emphasized by the results of experiment 2. Figure 3 shows three-point psychometric functions for the same three-component harmonic complex used in experiment 1, and for that complex with the center component mistuned. At all modulator delays, d' decreased monotonically with increases in mistuning, and was below one at a mistuning of 100 Hz in each direction.[3] Performance at the 100-Hz mistuning was significantly better than chance, but it will be argued in Sec. V B. that this was due to listeners using a weak harmonicity cue. For listener RC, performance with the perfectly harmonic complex was better than that for the identical stimulus in experiment 1. This may have been due to the fewer points measured in experiment 2, which would have reduced the signal uncertainty. Although a simple learning effect cannot be ruled out, there was no evidence for an improvement in performance during the course of either experiment.

Experiment 3 investigated the detection of FM incoherence at low carrier frequencies, where larger modulation depths are required in order for the FM to be detectable (Demany and Semal, 1989; Carlyon and Stubbs, 1989). Component frequencies were between 400 and 800 Hz, and the modulation depth was 5%. The results, plotted in Fig. 4, show that for two-component complexes with frequency ratios 2/3 (400 + *600* Hz) or 1/2 (400 + *800* Hz) listeners obtained very good performance, but that with an intermedi-

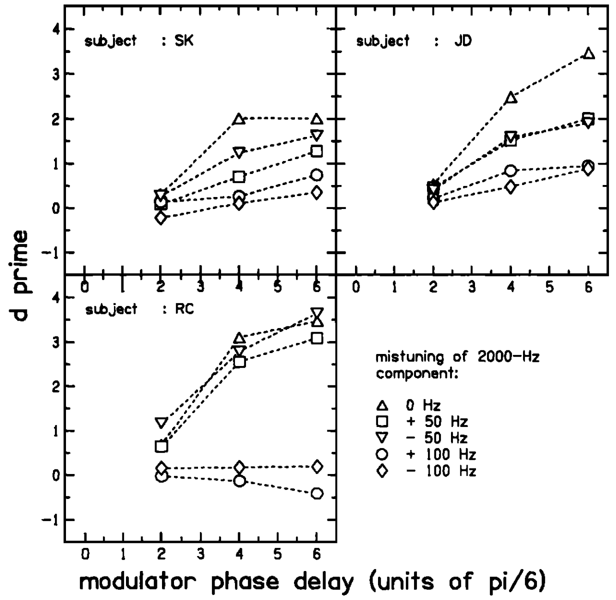

FIG. 3. As Fig. 2, for the five complexes of experiment 2. The five complexes differ in the degree of mistuning of the center component of a 1500 + 2000 + 2500-Hz complex.

ate frequency ratio of 4/7 (400 + *700* Hz) performance was much worse, and only JD performed significantly better than chance. In this experiment, it was particularly important to choose the inharmonic frequency ratio very carefully, in order to prevent listeners from detecting the target component moving in and out of tune with the 400-Hz component: pilot experiments revealed that choosing a target frequency even slightly different from 700 Hz led to a marked increase in performance.

The results of experiments 1–3 show that listeners can detect FM incoherence in sounds that are initially harmonic, but that detection becomes progressively worse as the sound is made less harmonic. This is true for component frequencies between 400–800 Hz, and for those between 1500–2500 Hz. It is suggested that listeners detect incoherence in harmonic sounds by virtue of the mistuning that it causes.

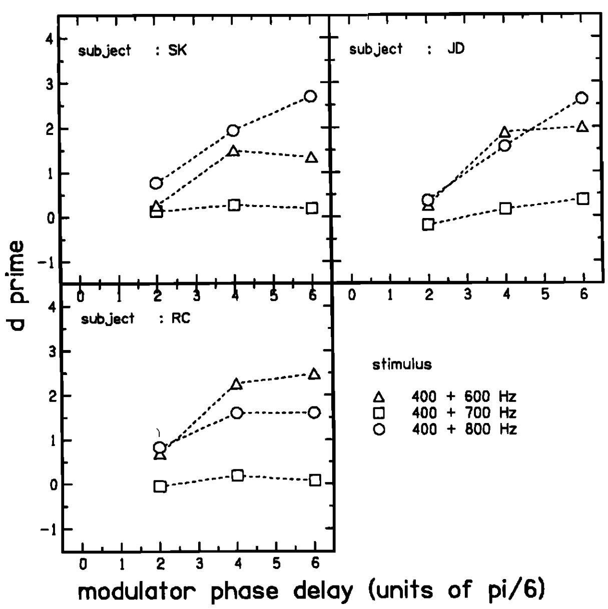

FIG. 4. As Fig. 2, for the three complexes of experiment 3.

II. EXPERIMENT 4: SENSITIVITY TO MISTUNING AND INCOHERENCE COMPARED

Listeners' insensitivity to FM incoherence in inharmonic sounds may mean that they are insensitive to FM incoherence *per se*, and that incoherently modulating one component of a harmonic complex sound is perceptually equivalent to imposing a simple mistuning on it. However, at least two alternative explanations are possible. First, many of the sounds that listeners might want to group on the basis of FM coherence are harmonic. A good example comes from the voiced portions of speech, where a periodic source undergoes FM, with all the harmonics of the fundamental being modulated coherently. The auditory system might be configured to detect FM coherence only for harmonic sounds. Second, the system might be more sensitive to the dynamic mistuning that is caused by incoherent FM than to a static mistuning. Therefore, experiment 4 compared psychometric functions for the detection of FM coherence and of a simple mistuning of the center component of a 1500 + *2000* + 2500-Hz complex. If there existed a separate mechanism for the detection of FM incoherence (over and above the mistuning that it caused), one might expect detection of incoherence to be superior to that for the detection of mistuning. If the same mechanism were involved in the two tasks, and if sensitivity to dynamic and to static mistunings varied similarly with the amount of mistuning, then the psychometric functions in the mistuning and incoherence conditions should be similar.

A. Mistuning condition

The stimuli were the same as in experiment 1, except that the target component (2000 Hz) was always modulated coherently but was mistuned in the signal interval by up to 100 Hz [see Fig. 1(b)]. Six mistunings were used, increasing in multiples of $\sqrt{2}$ from 18 to 100 Hz. The mistuning was produced by an increase in frequency for half the signal intervals and a decrease for the other half. (cf., Moore *et al.*, 1985). All modulations started at a positive-going zero crossing of the modulating waveform. In order to inhibit identification of the signal from the absolute frequency of the target component, the playback rate (and hence the $F0$) was randomized over $a + / - 5\%$ range from presentation to presentation (Carlyon and Stubbs, 1989). The cutoff frequencies of the reconstruction filters were also varied from presentation to presentation so as to always be 43% of the playback rate. In both this and the incoherence condition two modulation frequencies, 2.5 and 5 Hz, were used, and the signal duration was 400 ms.

B. Incoherence condition

The stimulus was the 1500 + *2000* + 2500-Hz complex of experiments 1 and 2 [see Fig. 1(a)], except that the playback rate was randomized for compatibility with the mistuning condition. Demany and Semal (1988) have shown that placing a modulator phase delay on one component relative to another causes their frequency relationship to vary quasi-sinusoidally between maxima and minima that depend on the value of the phase delay. The phase delays used here were chosen to produce maximum mistunings equal to those in the mistuning condition. These were, in units of π (mistunings in Hz in parentheses): 0.113 (18), 0.161 (25), 0.228 (35), 0.333 (50), 0.502 (71), and 1.0 (100). As the maximum mistunings increased in steps of $\sqrt{2}$, each modulator delay could also be compared to a mistuning equal to the rms mistuning that it produced (e.g., a phase delay of 0.228π produced a maximum mistuning of 35 Hz and an rms mistuning of 25 Hz).

C. Results and discussion

Figure 5 shows psychometric functions for the mistuning (triangles) and incoherence (squares) conditions at two modulation frequencies: 2.5 Hz (solid lines, shifted rightwards), and 5 Hz (dashed lines) (The symbols joined by dotted lines on the left-hand side of the data panel for listener JD, will be discussed later on). At both modulation frequencies, listeners were nearly always better at detecting a mistuning than at detecting an incoherence leading to an equivalent maximum mistuning (EMM). The only exceptions are for listener JD at $f_m = 5$ Hz and a mistuning of 100 Hz, and at $f_m = 2.5$ Hz and a mistuning of 71 Hz. However, when the comparison is made between detecting incoherence and its equivalent rms mistuning, performance in the incoherence condition is sometimes better than that in the mistuning condition. Thus, although listeners may have detected incoherence from the mistuning that it caused, they could not have done so by simply averaging over the entire waveform. Rather, they would have had sometimes to selectively weight those portions of the waveform during which mistuning was greatest.

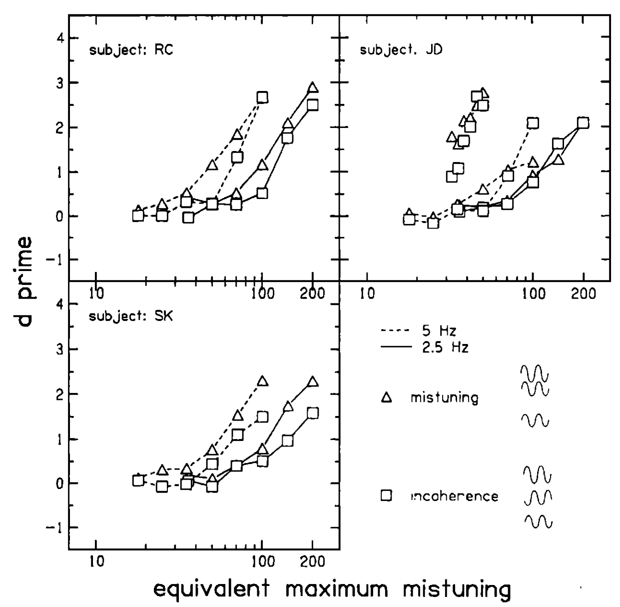

FIG. 5. Experiment 4. Detectability (d') of a mistuning (triangles) and a modulator phase delay (squares) of the center component of a 1500 + 2000 + 2500-Hz complex as a function of the "equivalent maximum mistuning" (see text for definition). Data obtained with $f_m = 5$ Hz are shown by dashed lines. Data obtained with $f_m = 2.5$ Hz are shown by solid lines and are shifted to the right for clarity. For listener JD, the points joined by dotted lines were obtained with $f_m = 5$ Hz for six closely spaced EMMs between 65 and 100 Hz, and are displaced to the left for clarity.

The motivation for experiment 4 was to compare the psychometric functions in the mistuning and incoherence conditions. Accordingly, linear least-squares fits were made to the functions relating d' to the logarithm of the EMM for each condition, with values of d' less than 0.2 excluded from the analysis (consistent with Carlyon *et al.*, 1990). The slopes of the fitted functions were compared using a t test (e.g., Edwards, 1973, pp. 187–189) and are shown in Table II. For an f_m of 2.5 Hz, there were no significant differences between the slopes in the two conditions. For an f_m of 5 Hz, the slopes for listener SK were not significantly different in the two conditions, but both RC and JD showed significantly steeper slopes in the incoherence than in the mistuning condition ($p < 0.05$ in each case). As these straight-line fits were based on a small number of points, the psychometric functions for $f_m = 5$ Hz were repeated for listener JD with more closely spaced EMMs (65, 71, 77, 84, 92, and 100 Hz). The resulting functions are plotted as the dotted lines in Fig. 5, and are offset leftward by two points. They confirm the initial findings in being steeper for the incoherence condition (slope = 10.06) than in the mistuning condition (slope = 5.78, $t = 2.47$, $df = 8$, $p < 0.05$). The new functions for JD show better performance at EMMs of 71 and 100 Hz than for the identical stimuli in her original functions for $f_m = 5$ Hz. This may have been due to the absence of very small EMMs in the new trials, which would have helped her maintain a better internal representation of the signal. The improvement might have been due to a learning effect, but there was no evidence of any improvement in performance during the measurement of the new function.

The functions for RC and JD at $f_m = 5$ Hz in the incoherence condition are steeper not only than those in the mistuning condition, but also than the functions obtained in the incoherence condition at $f_m = 2.5$ Hz. The reason for the slope differences is that the values of d' which fall just above the rejection criterion (0.2) in the incoherence condition at $f_m = 5$ Hz are smaller both than at $f_m = 2.5$ Hz in the same condition, and than in the mistuning condition at the same

TABLE II. Slopes of the psychometric functions for the "mistuning" and "incoherence" conditions of experiment 4, together with the results of t tests of the difference between the slopes in the two conditions [the asterisk (*) indicates $p < 0.05$]. Data for two modulation frequencies, 2.5 and 5.0 Hz, are shown. Also shown are "thresholds" (% rms mistuning corresponding to a d' of 0.78), obtained from the least-squares fits to the data in the "mistuning" condition.

	RC		JD		SK	
	2.5	5.0	2.5	5.0	2.5	5.0
Slopes						
Mistuning	4.51	4.06	3.12	2.17	4.34	3.44
Incoherence	5.22	7.96	4.14	7.95	2.62	3.51
t	0.76(ns)	4.24*	1.61(ns)	8.19*	−2.30(ns)	0.06(ns)
df	5	4	5	2	4	4
Thresholds						
(% mistuning)	1.3	1.3	1.5	2.0	1.6	1.4

f_m. It is therefore unlikely that RC and JD's steeper slopes in the incoherence condition at $f_m = 5$ Hz were due to the existence of an additional cue, specific to the detection of incoherence, that they could detect at large EMMs. Rather, the steep slopes seem to reflect an inability to detect efficiently the small dynamic mistunings caused by incoherent modulation at small EMMs.

The similarity between the psychometric functions in the mistuning and incoherence conditions at $f_m = 2.5$ Hz, together with the general superiority in the mistuning condition, suggests that listeners were not using a separate mechanism to detect FM coherence over and above that used for detecting mistuning. Thus the results of experiment 4 show no evidence for a mechanism specific to the detection of FM incoherence, independent of that used to detect mistuning.

III. EXPERIMENT 5: DETECTION OF MODULATOR PHASE DELAYS AS A FUNCTION OF MODULATION RATE

Experiments 1–3 showed that, for a modulation rate of 5 Hz and a signal duration of 400 ms, listeners do not detect modulation incoherence independently of the mistuning that it causes. However, listeners might be able to do so at much slower modulation rates and/or longer signal durations. It is known that the detection of linear frequency glides deteriorates with increases in glide rate (Dooley and Moore, 1988; Elliott *et al.*, 1989). This deterioration has been attributed to listeners having to take samples of the signal which are either shorter than, or whose mean frequency deviation from the standard is less than, those that could be obtained at slow modulation rates. To see whether similar limitations affect the detection of FM incoherence, experiment 5 measured psychometric functions for the detection of modulation incoherence as a function of modulation rate. The signal duration was increased to 1 s.

A. Stimuli and procedure

Stimuli and procedure were the same as for experiment 1, except as follows: listener SK was replaced with a new listener, HC, and there were only 100 trials per datapoint. The method of randomizing the signal starting phase was changed. A single 2-s sample of each stimulus was recorded on disk, and, for each interval of each trial, a 1-s sample was selected, starting at a point chosen at random between 0 and 1 s after the beginning of the waveform file. In experiment 5a, the signal was a 1-s version of the 1600 + *2300*-Hz signal of experiment 1, modulated by a 1-, 3-, 5-, or (for listener HC) 7-Hz sinusoid. In experiment 5b, the modulation rate was 1 Hz and three additional components were added to the stimulus of experiment 5a. The component center frequencies were either 767 + *1600* + 2300 + 3067 + 3833 Hz. (quasi-harmonic condition), or 774 + 1113 + *1600* + 2300 + 3306 Hz (inharmonic condition). The component frequencies in the quasiharmonic condition were chosen so as to help the listener detect the 1600-Hz component moving in and out of a 2/3 harmonic ratio with the 2300-Hz component: The ratios of the added component center frequencies to 2300 Hz were 1/3, 4/3, and 5/3. Thus all components except the target component (1600 Hz) formed a harmonic series with an $F0$ modulated around 767 Hz, and the incoherent modulation of the target component caused it to pass through the frequency of the second harmonic. The component frequencies in the inharmonic condition were chosen so as to make detection of changes in harmonicity more difficult, and were equally spaced on a logarithmic scale. Note that for both the harmonic and inharmonic complexes, the added component frequencies were sufficiently remote from the target frequency (1600 Hz) to produce negligible masking at that frequency relative to that produced by the pink noise background (Fletcher, 1940; Egan and Hake, 1950).

B. Results

The results of experiment 5a are shown by the dashed lines of Fig. 6. The two listeners who also took part in experiment 1, JD and RC, showed a small improvement at a modulation rate of 5 Hz, relative to their performance with the 400-ms stimuli of experiment 1. However, performance was much worse than for the harmonic stimuli of experiment 1, and d' was less than one at all modulator delays. All listeners performed better with at least one of the two slower modulation rates (1 and 3 Hz) than at 5 Hz, and listeners HC and JD obtained values of d' greater than two at the larger modulator delays. HC obtained a d' of 1.8 even at a modulation rate of 5 Hz, but this dropped to less than one when the modulation rate was increased to 7 Hz.

The results obtained in experiment 5a suggest that listeners can detect modulation incoherence for an inharmonic complex, but can only do so effectively when the modulation rate is slower than 5 Hz. One interpretation of this result is that there exists an across-frequency mechanism for detecting modulation incoherence, which operates only at low modulation rates and which is independent of the mechanism for detecting mistuning. An alternative explanation arises from the fact that the 1600 + *2300*-Hz "inharmonic" components did in fact move in and out of a 2/3 frequency

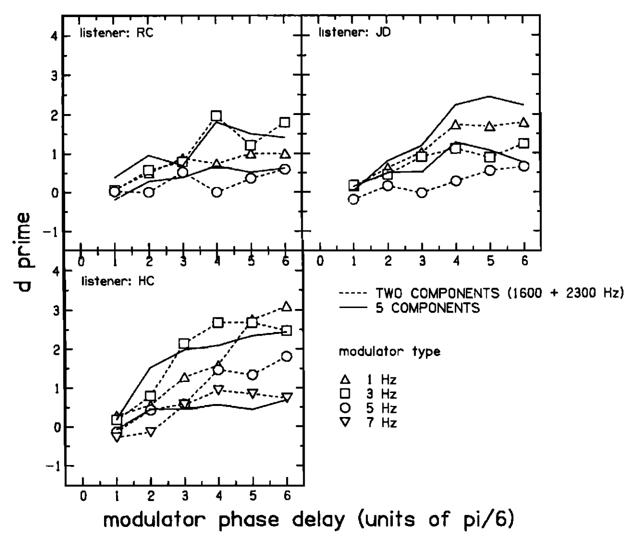

FIG. 6. As Fig. 2, for the data of experiments 5a and 5b. The dashed lines show data for experiment 5a as a function of f_m. The upper and lower solid lines in each panel are for the "quasi-harmonic" and "inharmonic" conditions, respectively, of experiment 5b.

ratio when modulated incoherently. This mistuning cue may have been too weak to detect at $f_m = 5$ Hz, but may have been detectable at lower f_m's. Experiment 5b tested these two explanations. If the first were true, then increasing the number of nontarget components, regardless of their frequency relationship to the original components, should either improve performance, or not affect it. This is because the listener would either be able to compare the output of the auditory filter responding to the target component with that of many other filters or, alternatively, would attend only to filters responding to two components. If the second explanation were correct, then adding nonharmonically related components should reduce performance, as the complex will sound mistuned throughout both signal and nonsignal intervals. The second explanation might also predict an increase in performance when components which form a harmonic series with the 2300-Hz component are modulated coherently with it.

The results of experiment 5b favor the "mistuning" explanation. The lower solid lines in Fig. 6 are psychometric functions for the (inharmonic) 774 + 1113 + *1600* + 2300 + 3306-Hz complex (modulation rate = 1 Hz). These functions all fall well below the corresponding 1-Hz functions for the two-component stimulus (dashed lines, triangles). In contrast, when components are added that correspond to a harmonic series based on a 767-Hz fundamental (767 + *1600* + 2300 + 3067 + 3833 Hz; upper solid line), performance for all listeners improves relative to that for the two-component stimulus at most modulator delays. The improvement in performance can be attributed to listeners hearing the 1600-Hz component moving in and out of tune with the harmonics of 767 Hz.

IV. EXPERIMENT 6: DETECTION OF INCOHERENCE CAUSED BY DIFFERENCES IN MODULATION RATE

Experiments 1–5 investigated the detection of modulation incoherence caused by differences in the phase of the waveforms modulating the individual components of complex sounds. Other investigators, too, have concentrated on modulator phase delays when investigating the processing of modulation incoherence (Schooneveldt and Moore, 1988; Gardner *et al.*, 1989). However, when two sounds are modulated by different waveforms the modulators usually differ in rate as well as in phase. Experiment 6 investigated the detection of modulation incoherence caused by differences in modulator rate.

A. Stimuli and procedure

Both harmonic (1500 + *2000* + 2500 Hz) and inharmonic (1500 + *1900* + 2500 Hz) carriers were used, and the signal duration was 400 ms. Modulators were sinusoids with frequencies of 2.5 Hz ("A"), 5 Hz ("B"), or 7.5 Hz ("C"), corresponding to one, two, and three modulation cycles per 400-ms stimulus, respectively. Three conditions A/B, A/C, and B/C, were used, corresponding to the two different modulation rates imposed on the different components. An example of the four possible trial structures is shown in Fig. 7, for condition B/C (f_m = 5.0 and 7.5 Hz). The standard always consisted of all three components modulated at the same rate: the lower (L) of the two possible f_m's on half the trials and the higher (H) f_m on the other half. (See the "standard" stimuli in the left- and right-hand parts of Fig. 7, respectively.) On half the trials the signal consisted of the highest and lowest carrier frequencies being modulated by L and the center component by H [signal = "low-high-low (LHL)," top of Fig. 7]; on the other half the highest and lowest components were modulated by H and the center component by L [signal = "high-low-high (HLH)," top of Fig. 7]. For each condition the standard and signal waveforms were chosen at random for each 2I, 2AFC trial. This was done to prevent listeners from detecting the signal on the basis of the modulation frequency imposed on any one component: the listener was required to compare the f_m's across the different carrier components, and to make a decision based on whether or not they were the same. A "phase delay" condition, $P0/P\pi$, was also included, so that the detection of incoherence due to differences in f_m and to differences in phase could be compared using the same paradigm. In condition $P0/P\pi$, the modulator frequency was always 5 Hz, and the waveform modulating each component started at either a negative ("0") or positive ("π") zero crossing. For the standards, the waveforms modulating each component were in the same phase, whereas for the signals the modulation of the center component was inverted relative to that of the other two components. For all conditions, data were collected in ten trial blocks, each consisting of ten repetitions of each of the four possible combinations of standard and signal. However, listener SK dropped out toward the end of the study, and the analyses for his data in conditions A/B, B/C, and A/C are based on only eight blocks.

For conditions A/B, A/C, and B/C, each component was modulated by 3.54% of its starting frequency. This was done so that, for the harmonic stimuli, the rms mistuning would be equal to that obtained with a π phase delay and a

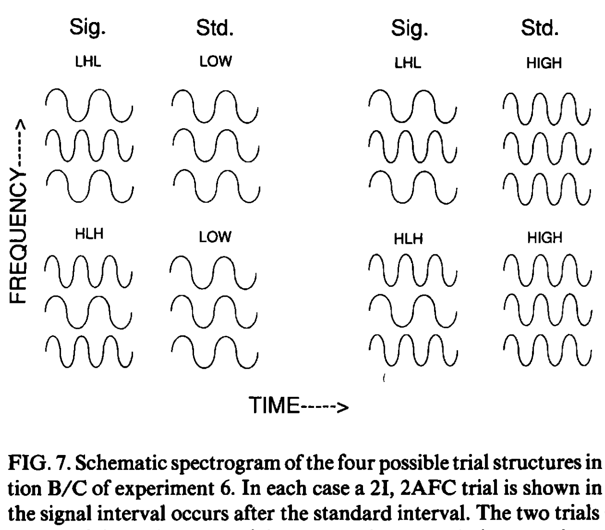

FIG. 7. Schematic spectrogram of the four possible trial structures in condition B/C of experiment 6. In each case a 2I, 2AFC trial is shown in which the signal interval occurs after the standard interval. The two trials shown on the left have the lower of the two possible modulation rates imposed on all components of the standard; the trials on the right have the higher modulation rate imposed on the standard. The signals in the bottom row have the end components modulated at the higher rate and the center component modulated at the lower rate (signal = HLH); the converse is true for the signals (LHL) in the top row.

2.5% modulation depth in experiment 1 (footnote 4). In condition $P0/P\pi0$, the modulation depth was 2.5%

B. Results

The results of experiment 6 are shown as % correct scores for each listener and masker/signal combination in Table III. Scores that are shown in bold type were significantly greater than chance; those shown in italics were less than chance ($p<0.05$, *post hoc* test[5]). For the harmonic stimulus, listeners RC and JD performed above chance for all conditions. Listener SK's performance was mostly above chance, but not always significantly so. For the inharmonic stimulus, the pattern of results was more complex, and varied across listeners. Therefore, a separate analysis of variance (modulator pair × harmonicity × standard × signal; 3 × 2 × 2 × 2) was performed on each listener's data for the modulator pairs A/B, A/C, and B/C.[6] The data in condition $P0/P\pi$ were analyzed separately.

TABLE III. Mean scores (percent correct) for each condition of experiment 6. The top row of scores in each condition is for listener RC, the middle row is for JD, and the bottom for SK. In conditions A/B, A/C, and B/C the two left-hand columns are for conditions in which all components of the standard were modulated at the lower of the two possible f_m: the signal had either the top and bottom components modulated at the lower f_m and the center component modulated at the higher f_m ("LHL"), or vice versa ("HLH"). The two right-most columns are for conditions in which the standard had all components modulated at the higher of the two f_m (A = 2.5 Hz, B = 5.0 Hz, C = 7.5 Hz). In condition $P0/P\pi$, the left- and right-hand columns are for standards whose modulation started at negative and positive zero crossings, respectively. Scores that are significantly greater than chance [50%, $p<0.5$ (Ref. 5)] are in bold type; those significantly worse than chance are in italics.

Reference f_m	Low		High	
Signal	LHL	HLH	LHL	HLH
(a) Harmonic				
	87	**88**	**77**	**88**
A/B	**95**	**99**	**89**	**90**
	59	71	57	57
	83	**91**	67	**81**
A/C	**89**	**98**	**86**	**95**
	62	46	52	53
	80	**90**	*33*	65
B/C	**81**	**81**	**81**	**76**
	63	48	60	44
	82	**79**	**73**	**82**
$P0/P\pi$	**97**	**98**	**94**	**97**
	62	61	67	58
(b) Inharmonic				
	65	66	*26*	55
A/B	**78**	**82**	57	62
	54	57	51	43
	72	**93**	36	64
A/C	**71**	60	**88**	**86**
	67	66	33	46
	68	**80**	*32*	52
B/C	63	52	**71**	67
	53	44	54	51
	51	**71**	37	51
$P0/P\pi$	**68**	**77**	57	67
	52	52	44	51

1. Incoherence in modulation rate

As can be anticipated from the data in Table III, there was a significant main effect of harmonicity for all three listeners ($p<0.01$ in each case). Table III shows that performance was above chance in some of the inharmonic conditions. For listener RC, this can best be explained by a tendency to identify any stimulus containing the higher of the two possible modulation rates as the signal (i.e., as being less "coherent"). Accordingly, he performed significantly better when the standard contained the lower modulation frequency (two left-hand columns of Table III) than when it contained the higher (right-hand columns), and this difference was greatest for inharmonic stimuli (main effect of standard, $p<0.01$; standard × inharmonicity, $p<0.01$). When both the signal and standard contained the higher f_m (two right-hand columns), his average discrimination of the inharmonic stimuli was not better than chance.

There is evidence that listener JD's responses were also influenced by overall modulation rate in some conditions. She too showed an advantage for standards modulated at the lower of the two possible rates for the modulator pair A/B, but showed the opposite effect for pairs A/C and B/C, at least for the inharmonic stimuli. This led to a significant interaction between modulator pair and standard type ($p<0.01$), and a significant three-way interaction between modulator pair, standard, and harmonicity ($p<0.05$). The data for RC and JD indicate that, although listeners can perform above chance with some inharmonic stimuli by adopting certain strategies, they perform worse and less consistently than when strong harmonicity cues exist. Listener SK did not perform significantly above chance with inharmonic stimuli, presumably because he did not learn to adopt these strategies. It is unclear which listener's data most accurately reflect the strategies used in everyday situations.

2. Incoherence in modulator phase

For the $P0/P\pi$ condition an analysis of variance (harmonicity × standard × signal: 2 × 2 × 2) revealed significant effects of harmonicity for all three listeners ($p<0.01$ for RC and JD, $p<0.05$ for SK). This confirms the results of experiments 1–3. The only other significant effect was an interaction between harmonicity and signal type for listener RC ($p<0.05$).

Listeners' performance in the present paradigm may be compared to that observed in experiment 2 by averaging the scores in the harmonic and inharmonic conditions separately, and converting them to d' under the assumption of no response bias. Such a comparison yields comparable values in the inharmonic condition of experiment 6 (RC:0.2, JD:0.6, SK:0.0) as in experiment 2 (0.2, 0.9, and 0.3, respectively). In the harmonic condition values of d' were lower in experiment 6 (1.1, 2.7, 0.4) than in experiment 2 (3.5, 3.5, 2.0). The reasons for this are not obvious, although they are probably related to the intermingling of the phase delay conditions with the f_m incoherence conditions. As described above, listeners had a tendency to listen for gross changes in modulation rate, and this may have depressed performance in the $P0/P\pi$ condition.

V. DISCUSSION

A. Comparison with previous studies

1. Detecting incoherence

The experiments presented here investigated the detection of an across-frequency difference in the modulations imposed on the different components of a complex sound. The results show that, when modulation incoherence causes one component to become mistuned from an otherwise harmonic complex, listeners obtain near-perfect performance, but that when this "mistuning" cue is unavailable performance is greatly reduced. The finding is qualitatively similar to that reported by McAdams (1984), and by Demany and Semal (1988). However, both the previous studies reported detection of FM incoherence for inharmonic complexes that was superior to that shown here.

McAdams presented listeners with 16-component complexes that were either coherently modulated or had one component modulated independently of the others. He measured discriminability (the proportion of trials on which the incoherently modulated stimulus was judged as "having more sound sources in it"), as a function of modulation depth, and found superior performance for harmonic than for inharmonic stimuli. It is likely that in the harmonic condition, his listeners were detecting inharmonicity, rather than incoherence *per se*. However, his listeners could also achieve near-perfect performance for the inharmonic stimuli, given a sufficiently large modulation depth (typically between 6 and 8 cents, or 0.35%–0.46% rms). The superior performance in the inharmonic condition of McAdams' study over that reported here may be attributed to two differences between the stimuli used. First, McAdams stressed that the components of his inharmonic stimuli were mistuned only slightly from harmonic frequencies: the mistuning ranged from 0–5 cents (0%–0.29%) and the unmodulated waveform contained "a vague quasi-periodicity of about the same period as the harmonic waveform." As his listeners could detect incoherence only at modulation depths greater than this mistuning, it is possible that they did so by detecting the increase in mistuning caused by incoherently modulating one component. Second, McAdams' stimuli were presented at a level of 75 dBA in quiet, whereas ours were presented at a lower level (45 dB SPL/component) in a background of pink noise. It is possible that his listeners made use of within-channel cues, such as changes in the rate of beating between adjacent, incoherently modulated components. Such within-channel cues were observed when some of the stimuli (e.g., 1500 + *2100* + 2500 Hz) of the present experiments were presented informally in the absence of noise.

Demany and Semal (1988) used an adaptive procedure, and required listeners to discriminate between coherently and incoherently modulated two-tone complexes. Because their stimuli were presented dichotically, within-channel mechanisms cannot have influenced their results. They found that listeners could detect a smaller modulator phase delay between components that were harmonically related (or nearly so) than between mistuned components, but that listeners could reliably detect a modulator phase delay of $\pi/6 - \pi/3$ ($d' = 1.13$) for components mistuned from an octave relationship. The largest mistuning that they used was 100 cents (5.7%), which is slightly greater than the 5% mistuning of the 1500 + *1900* + 2500-Hz complex of experiment 2, for which the maximum d' obtained by any of our listeners was less than one. However, Demany and Semal used a 5% (0-peak) modulation depth, twice that used in experiment 2. It is likely that this allowed their listeners to detect the target component moving in and out of tune with the remaining component. Their slower (2-Hz) modulation rate might also account for the superior performance of their listeners: Experiment 5 showed that when f_m is between 1 and 3 Hz, listeners can use weak mistuning cues that are unavailable when $f_m = 5$ Hz.

2. Detecting mistuning

Experiment 4 measured psychometric functions for the detection of mistuning of the center component of a frequency-modulated three-component complex. The paradigm is similar to that used by Moore *et a*l. (1985), the main differences being that their stimuli were unmodulated ten- or 12-component complexes, and that they used an adaptive procedure. They reported a mean threshold ($d' = 0.78$) for the fourth harmonic of a 400-Hz $F0$ of about 1% of the target frequency. Table II shows that thresholds ($d' = 0.78$), expressed in rms mistuning and estimated from the least-squares fits to the data of experiment 4, range from 1.3%–2% of the target frequency (2000 Hz) and are slightly higher than that reported by Moore *et al.* (1985). One reason for the superior performance reported by Moore *et al.* (1985) is that they used a larger number of components than reported here: experiment 5b of the present study showed that adding additional harmonically spaced components to a two-component complex leads to an increase in sensitivity, presumably by causing an increase in the pitch strength of the reference stimulus.

B. Detection of incoherence with inharmonic complexes

Detection of FM incoherence in the present study was greatly reduced for inharmonic complexes compared to harmonic complexes, and was usually so bad that a threshold could not have been measured had an adaptive procedure (e.g., Levitt, 1971) been used. In several conditions (e.g., the inharmonic condition of experiment 1), performance was not significantly different from chance. Nevertheless, in some experiments (e.g., 2 and 5a), performance was above chance even for inharmonic complexes, and there are two alternative explanations for this.

One possibility is that listeners *can* detect FM coherence *per se*, but that the mechanism is not very sensitive, at least for the stimuli used here. This could be because the modulation characteristics used here were different from those which listeners normally encounter, or that the complexes had too few components whose modulations could be compared. Horii (1989) measured the periodic frequency modulation (vibrato) that occurs in singing and reported a mean frequency of 5.7 Hz and a mean zero-peak excursion of 3.4%. These values are close to those used in the present study (2.5 Hz and 5 Hz; 2.5% and 5%). The modulations used here can also be compared to those occurring in speech. $F0$ changes in speech are rarely periodic, so it is more instructive to consider rates of frequency change. For our 5 Hz, 2.5% (zero-peak) FM, each component changed frequency by 5% over the 100-ms half-cycle of modulation from $\pi/2$ to $3\pi/2$. Lieberman (1967) presented $F0$ contours of normal speech with $F0$'s changing by 5% in as little as 10 ms (e.g., his Fig. 7). He also presented contours that remained almost constant over 100-ms intervals. It seems, therefore, that the stimuli used here had rates of change not atypical of those encountered in either speech or music. The present stimuli were perhaps a little artificial in that they had only a few components, and in that the task was to "hear out" only a single component. However, it is worth noting that the auditory system is capable of perceptually separating a single component on the basis of other cues such as mistuning (with the present stimuli) and onset asynchrony (Viemeister, 1980; Viemeister and Bacon, 1982; Carlyon, 1989)

The results of an experiment using more than three components gives rise to a second explanation for listeners' above-chance performance at detecting FM incoherence in some inharmonic complexes. Experiment 5 compared the detection of incoherence in a two-component (1600 + 2300 Hz) complex, with that of the same complex with additional logarithmically spaced components modulated coherently with the 2300 Hz component. The reduction in performance caused by the added components argues against the detection of incoherence *per se*, and in favor of the detection of a weak harmonicity cue. The inharmonic stimuli of experiment 2, for example, were fairly close to being harmonic, and listeners could have detected the target component moving in and out of a harmonic relationship with the other components. This would have been possible for even the greatest mistuning (100 Hz) of the center component: For the 1500 + *1900* + 2500-Hz complex, the component frequency ratios were 3.0/3.8/5.0 at a zero modulator delay, but momentarily reached the harmonic ratio 3.0/4.0/5.0 when the modulator phase delay equaled π. Such a strategy would have been even more useful when the mistuning was only 50 Hz. It is likely that if additional logarithmically spaced components had been added to the mistuned stimuli, performance would have been even worse relative to that obtained with harmonic stimuli.

C. Relevance to the perceptual separation of concurrent sounds

The results of the present study suggest that there exists no across-frequency mechanism specific to the detection of FM coherence. The data also show that listeners can detect FM incoherence when it results in mistuning of the target component; there is also evidence that they can do so on the basis of FM-induced AM (Wakefield and Edwards, 1989) and when the incoherence results in beating between adjacent components (informal observation described in Sec. V A 1). It is therefore worth drawing a distinction between the absence of a mechanism specific to the detection of FM coherence, and the fact that listeners may detect FM coherence indirectly on the basis of a covarying cue. There is evidence that listeners can use at least one such covarying cue, mistuning, to perceptually separate concurrent complex sounds (Brokx and Nooteboom, 1982; Scheffers, 1983; Summerfield and Assmann, 1990). When such additional cues are not present, listeners cannot detect incoherence in a frequency modulation that is easily detectable, sufficient for the detection of mistuning, and typical of that encountered in music and speech.

ACKNOWLEDGMENTS

This research was supported by a project grant from the Medical Research Council (U.K.), and by a Royal Society University Research Fellowship to the author. I would like to thank C. J. Darwin, B. C. J. Moore, and L. Demany for useful comments on a previous version of this article. Mark Lutman provided some low-level software.

[1]Although a 3I,2AFC procedure was used here, d' was calculated as if a 2I, 2IFC procedure had been used. In other words, it was assumed that the listener compared the second and third intervals in each trial and decided which was the "more incoherent." The purpose of the first interval was simply to remind the listener of the reference stimulus, and it was assumed to have no effect other than to make performance more similar to that of an ideal observer in a 2I,2IFC task. In justification of this method, it can be noted that in most commonly used procedures the reference stimulus is presented many times throughout a block of trials, and so the listener usually has some stored representation of it.

[2]Using a normal approximation to the binomial distribution, it can be shown that a score of 57% is required to reject ($p < 0.05$) the hypothesis that the probability of a correct response is equal to 0.5. A score of 57% corresponds to a d' of 0.25 in the two-alternative task used here.

[3]One possibility is that the reduced performance at the 100-Hz mistunings relative to that with no mistuning was due to the change in frequency spacing between adjacent components, rather than to the inharmonicity. Whereas the harmonic complex had spacings of 500 Hz, the mistuned complex (e.g., 1500 + *1900* + 2500 Hz) contained spacings of 400 and 600 Hz. Therefore an auxiliary experiment was performed with a harmonic complex containing 400-Hz spacings (1600 + *2000* + 2400 Hz), and with a complex containing 600-Hz spacings (1200 + *1800* + 2400 Hz). Listeners performed well above chance with both these complexes.

[4]Demany and Semal (1988) showed that when two components are frequency modulated at the same f_m but with a π modulation delay, their frequency ratio oscillates sinusoidally around the starting frequency ratio. Their analysis also demonstrated that the rms deviation from the original ratio is equal to the (zero-peak) modulation depth multiplied by $\sqrt{2}$, provided that the modulation depth is a small percentage of each component frequency. When the two components are modulated at different f_m's the variation in frequency ratio with time is more complex, and a formal analysis difficult. We approached the problem numerically using a computer simulation and found the rms deviation to be equal to the modulation depth. The solution is intuitively reasonable, as it yields the same rms deviation as when the modulators have the same frequency but are uncorrelated ($\pi/2$ delay) (Demany and Semal, 1988).

[5]The 95% confidence limits surrounding chance performance were calculated from the normal approximation to the binomial distribution for $N = 100$, the number of trials for each entry in the table. As there were 96

such entries, the criterion probability for significance was adjusted from 0.05 to the 96th root of $(1-0.05) = 0.000534$, yielding confidence limits from 33.7 to 67.3. The corresponding limits for listener SK in the "incoherence-by-rate" conditions, for which $N = 80$, are 30.7 and 69.3, respectively.

[6]Only those significant effects relevant to the discussion are described in the text. A complete list of significant main effects and interactions follows. RC: harmonicity ($p < 0.01$), modulator pair ($p < 0.01$), standard ($p < 0.01$), signal ($P < 0.01$), harmonicity × modulator pair ($p < 0.01$), signal × standard ($p < 0.01$), modulator pair × standard ($p < 0.05$), harmonicity × standard ($p < 0.05$). JD: harmonicity ($p < 0.01$), modulator pair ($p < 0.01$), modulator pair × standard ($p < 0.01$), modulator pair × signal ($p < 0.05$), modulator pair × standard × harmonicity ($p < 0.05$). SK: harmonicity ($p < 0.01$), standard ($p < 0.05$), modulator pair × standard ($p < 0.01$), modulator pair × standard × harmonicity ($p < 0.01$).

ANSI (1969). ANSI S3.6-1969, "Specifications for audiometers" (American National Standards Institute, New York).

Brokx, J. P. L., and Nooteboom, S. G. (1982). "Intonation and the perceptual separation of simultaneous voices," J. Phon. 10, 23–36.

Carlyon, R. P. (1989). "Changes in the masked thresholds of brief tones produced by prior bursts of noise," Hear. Res. 41, 223–236.

Carlyon, R. P., and Stubbs, R. J. (1989). "Detecting single-cycle frequency modulation imposed on sinusoidal, harmonic, and inharmonic carriers," J. Acoust. Soc. Am. 85, 2563–2574.

Carlyon, R. P., Buus, S., and Florentine, M. (1990). "Temporal integration of trains of tone pulses by normal and by cochlearly impaired listeners," J. Acoust. Soc. Am. 87, 260–268.

Chalikia, M. H., and Bregman, A. S. (1989). "The perceptual segregation of simultaneous auditory signals: pulse train segregation and vowel segregation," Percept. Psychophys. 46, 487–496.

Chalikia, M. H., and Bregman, A. S. (1990). "The perceptual segregation of simultaneous vowels with harmonic, shifted, and random components," submitted to Percept. Psychophys.

Cherry, E. C. (1953). "Some experiments on the recognition of speech, with one and two ears," J. Acoust. Soc. Am. 25, 975–979.

Darwin, C. J., and Gardner, R. B. (1986). "Mistuning a harmonic of a vowel: Grouping and phase effects in vowel quality," J. Acoust. Soc. Am. 79, 838–845.

Demany, L., and Semal, C. (1988). "Dichotic fusion of two tones one octave apart: Evidence for internal octave templates," J. Acoust. Soc. Am. 83, 687–695.

Demany, L., and Semal, C. (1989). "Detection thresholds for sinusoidal frequency modulation," J. Acoust. Soc. Am. 85, 1295–1301.

Dooley, G. J., and Moore, B. C. J. (1988). "Detection of linear frequency glides as a function of frequency and duration," J. Acoust. Soc. Am. 84, 2045–2057.

Edwards, A. E. (1973). *Statistical Methods* (Holt, Rhinehardt and Winston, New York), 3rd ed.

Elliott, L. L., Hammer, M. A., Scholl, M. E., Carrell, T. E., and Wasowicz, M. (1989). "Discrimination of rising and falling simulated single-formant frequency transitions: Practice and transition duration effects," J. Acoust. Soc. Am. 86, 945–953.

Egan, J. P., and Hake, H. W. (1950). "On the masking pattern of a simple auditory stimulus," J. Acoust. Soc. Am. 22, 622–630.

Fletcher, H. (1940). "Auditory patterns," Rev. Mod. Phys. 12, 47–65.

Gardner, R. G., and Darwin, C. J. (1986). "Grouping of vowel harmonics by frequency modulation: absence of effects on phonemic categorization," Percept. Psychophys 40, 183–187.

Gardner, R. G., Gaskill, S., and Darwin, C. J. (1989). "Perceptual grouping of formants with static and dynamic differences in fundamental frequency," J. Acoust. Soc. Am. 85, 1329–1237.

Grose, J. H., and Hall, J. W. (1990). "The effect of modulation coherence on signal threshold in frequency-modulated noise bands," J. Acoust. Soc. Am. 88, 703–710.

Hall, J. W., Haggard, M. P., and Fernandes, M. A. (1984). "Detection in noise by spectro-temporal pattern analysis," J. Acoust. Soc. Am. 76, 50–56.

Horii, Y. (1989). "Frequency modulation characteristics of sustained /a/ sung in vocal vibrato," J. Speech. Hear. Res. 32, 829–836.

Levitt, H. (1971). "Transformed up–down methods in psychophysics," J. Acoust. Soc. Am. 49, 467–477.

Lieberman, P. (1967). *Intonation, Perception, and Language*, Res. Monogr. No. 38 (MIT, Cambridge, MA).

McAdams, S. (1984). "Spectral fusion, spectral parsing and the formation of the auditory image," Ph. D thesis, Stanford University, Stanford, CA.

McAdams, S. (1989). "Segregation of concurrent sounds. I.: Effects of frequency modulation coherence " J. Acoust. Soc. Am. 86, 2148–2159.

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

Parsons, T. W. (1976). "Separation of speech from interfering speech by means of harmonic selection," J. Acoust. Soc. Am. 60, 911–918.

Schooneveldt, G. P., and Moore, B. C. J. (1987). "Comodulation masking release (CMR): Effects of signal frequency, flanking-band frequency, masker bandwidth, flanking-band level, and monotic versus dichotic presentation of the flanking band," J. Acoust. Soc. Am. 82, 1944–1956.

Schooneveldt, G. P., and Moore, B. C. J. (1988). "Failure to obtain comodulation masking release with frequency modulated maskers," J. Acoust. Soc. Am. 83, 2290–2292.

Scheffers, M. T. M. (1983). " Sifting vowels: auditory pitch analysis and sound segregation," Doctoral dissert., Univ. of Groningen, The Netherlands.

Summerfield, A. Q., and Assmann, P. P. (1990). "Perception of concurrent vowels: Effects of pitch-pulse asynchrony and harmonic misalignement," submitted to J. Acoust. Soc. Am.

Stubbs, R. J., and Summerfield, A. Q. (1988). "Evaluation of two voice-separation algorithms using normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. 84, 1236–1249.

Viemeister, N. F. (1980). "Adaptation of masking," in *Psychophysical, Physiological and Behavioral Studies in Hearing,* edited by G. van den Brink and F. A. Bilsen (Delft U.P., Delft, The Netherlands), pp. 190–198.

Viemeister, N. F., and Bacon, S. P. (1982). "Forward masking by enhanced components in harmonic complexes," J. Acoust. Soc. Am. 71, 1502–1507.

Wakefield, G. H., and Edwards, B. (1989). " Cross-spectral envelope phase discrimination for FM signals," J. Acoust. Soc. Am. Supp. 1 85, S122.