PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

The Psychophysics of Concurrent Sound Segregation [and Discussion]

Robert P. Carlyon and S. McAdams

Phil. Trans. R. Soc. Lond. B 1992 **336**, 347-355
doi: 10.1098/rstb.1992.0068

The psychophysics of concurrent sound segregation

ROBERT P. CARLYON

Laboratory of Experimental Psychology, University of Sussex, Brighton BN1 9QG, U.K.

SUMMARY

To perceptually separate concurrent complex sounds, normally hearing listeners simultaneously combine information across a wide range of frequency components. Three psychoacoustical experiments are described which investigate different forms of this across-frequency processing. The first two experiments investigate the role of coherence of frequency modulation (FM) between widely separated frequency components of a complex sound. The first experiment bolsters existing evidence that, for harmonic sounds, listeners can discriminate coherent from incoherent FM, but only by detecting the mistuning that arises from incoherent FM. The second demonstrates that, for inharmonic sounds, coherence of FM has no effect on the phenomenon of modulation detection interference (see Moore & Shailer, this symposium) once within-channel cues (combination tones and beating) are masked by background noise. It is concluded that there is not an across-frequency mechanism specific to the detection of FM incoherence. The third experiment investigates the extent to which the detection of mistuning of one component of a harmonic complex is impaired by an interfering sound (the 'interferer') with a frequency spectrum similar to that of the mistuned component. When the interferer is gated on and off with the harmonic complex, it has only a small effect provided that its level is more than 3 dB below that of the target. However, when the interferer starts before and ends after the complex, thresholds are elevated more, and this elevation occurs even for low-level interferers. Explanations of this effect in terms of adaptation and of auditory streaming are discussed.

1. INTRODUCTION

Until early last decade, most psychoacoustical research concentrated on the ability of listeners to make sequential comparisons between pairs of stimuli that differed along a single dimension, such as the intensity or frequency of a pure tone. Such experiments allowed the development of quite precise theories of basic auditory processes, and of models that could accurately predict experimental data. For example, the threshold for a pure tone signal in a masking band of noise can be accurately predicted from the amount of masker energy passing through an 'auditory filter' centered on the tone (Fletcher 1940; Zwicker *et al.* 1957; Patterson 1976). More recently, psychoacousticians have developed techniques for studying the processes used in many everyday listening situations, where we have to perform simultaneous comparisons of energy falling in different frequency regions. Across-frequency processing is important for at least two real-life tasks. First, to identify the spectral shape of an isolated sound (e.g. a vowel), one has to compare the levels of different frequency components. Perhaps more importantly, the perceptual separation of two sounds with overlapping spectra requires listeners to 'sort through' the combined spectrum, identifying which frequency components belong to which sound. The ability of listeners to perform both of these types of simultaneous, across-frequency comparisons, has been the subject of much recent experimental work (Hall *et al.* 1984; Moore & Glasberg 1986; Demany & Semal 1988; Green 1988; Carlyon & Stubbs 1989; Demany *et al.* 1991). The experiments reported here investigate two potential cues to concurrent sound segregation that require across-frequency processing.

(*a*) F_0 *differences*

Periodic sounds, such as the vowels of speech contain frequency components ('harmonics') equal to integer multiples of a common fundamental ('F_0'). Consequently, frequency components that do not correspond to one of these harmonics can be attributed to a different sound source, and can be discriminated from an in-tune harmonic, even when all harmonics excite separate auditory filters (Moore *et al.* 1985*b*; Moore & Glasberg 1986; Demany & Semal, 1988, 1991; Hartmann *et al.* 1990; Demany *et al.* 1991; Carlyon *et al.* 1992). It is also known that listeners can use F_0 differences between competing sources of voiced speech to improve the identification of the constituent sounds (Brokx & Nooteboom 1982; Scheffers 1983; Summerfield & Assmann 1991).

(*b*) *FM coherence*

A second, attractive, potential cue for the perceptual grouping of different components of a complex sound is FM coherence. This refers to the fact that

Phil. Trans. R. Soc. Lond. B (1992) **336,** 347–355
Printed in Great Britain

when the fundamental frequency (F_0) of, for example, a speaker's voice changes, all the frequency components of that voice change in the same direction at the same time. It is plausible that listeners group together the coherently changing components of a single voice, and separate them from components arising from a different source, which may be modulated incoherently with that voice. If so, the auditory system might use FM coherence in the same way that the visual system processes coherence of spatial movement to group different parts of the same object. By analogy with vision, we might imagine a mechanism whereby the auditory system not only identifies the peaks in the basilar membrane excitation pattern caused by individual components, but also correlates their movement along it (McAdams 1984; Bregman 1990; Wilson *et al.* 1990). However, it is difficult to show conclusively that FM incoherence per se can facilitate the perceptual separation of frequency components because, when a component of a harmonic complex is modulated incoherently from the others, it also becomes mistuned. Listeners might detect this mistuning, rather than the FM coherence per se.

The first experiment reported here extends previous work suggesting that, once within-channel cues are removed, listeners are not sensitive to FM incoherence, independently of the mistuning that it causes. The second experiment resolves a discrepancy between the results of experiment 1 and those of a study (Wilson *et al.* 1990) that investigated the influence of FM incoherence on the phenomenon of 'modulation detection interference' (see Moore & Shailer, this symposium). Finally, a third experiment measures thresholds for the detection of mistuning imposed on one component of a harmonic complex, both in the presence and absence of an interfering sound. The 'interferer' has a frequency spectrum similar to that of the potentially mistuned harmonic, and is either gated synchronously with the complex or starts before and ends after it.

2. DETECTION OF FM INCOHERENCE

(*a*) *Background*

In a recent article (Carlyon 1991), it was argued that listeners are not sensitive to FM coherence in the absence of additional cues such as mistuning. In one experiment, it was shown that although listeners could discriminate between coherent and incoherent sinusoidal FM of harmonic complexes, they could not do so when the complexes were inharmonic. A second experiment showed that listeners' psychometric functions for the detection of a static mistuning imposed on one component of a harmonic complex could account for the corresponding functions describing the detection of FM coherence. The article concluded that there is no across-frequency mechanism specific to the detection of FM incoherence.

The finding that listeners cannot discriminate coherent from incoherent FM, if generalizable to all stimuli, demonstrates that across-frequency comparisons of FM coherence cannot play a role in perceptual sound segregation. However, it is possible that the findings reported by Carlyon (1991) were specific to the sinusoidal nature of the FM used. As figure 1*a* shows, imposing incoherent sinusoidal FM on two sounds causes their F_0s to spend a moderate amount of time at their maximum mistuning. During this time the F_0s are roughly constant, and therefore do not change incoherently: thus, the choice of a sinusoidal pattern of modulation could have biased listeners away from detecting incoherence, and towards the detection of mistuning. The aim of the first experiment reported here was to test this explanation by comparing the detection of incoherence obtained with sinusoidal modulation with that obtained using triangular FM. As figure 1*b* shows, incoherent triangular FM results in the two F_0s reaching their maximum mistuning only momentarily, while spending virtually all their time moving in opposite directions. If listeners are sensitive to FM incoherence, then performance should be better with triangular than with sinusoidal FM. If, however, they are sensitive only to mistuning, then the opposite should be true. In addition, psychometric functions were measured for the detection of static mistunings, using the two types of modulation pattern. If listeners detect incoherence only from the mistuning that it causes, then, for a given modulation pattern, psychometric functions for the detection of incoherence and of static mistunings should be similar.

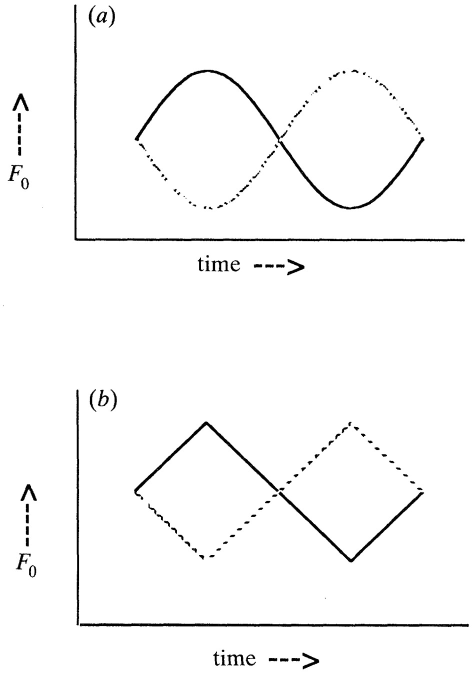

Figure 1. Schematic representation of the pattern of F_0 change imposed by two patterns of FM. (*a*) Sinusoidal FM; (*b*) triangular FM. The two traces in each part of the figure represent the F_0s of two sounds that are modulated incoherently (π modulator delay).

(*b*) *Stimuli and procedure*

The method of stimulus generation and procedure were as described by Carlyon (1991), except for the inclusion of a condition with triangular modulation. Briefly, a three-interval, two-alternative forced-choice

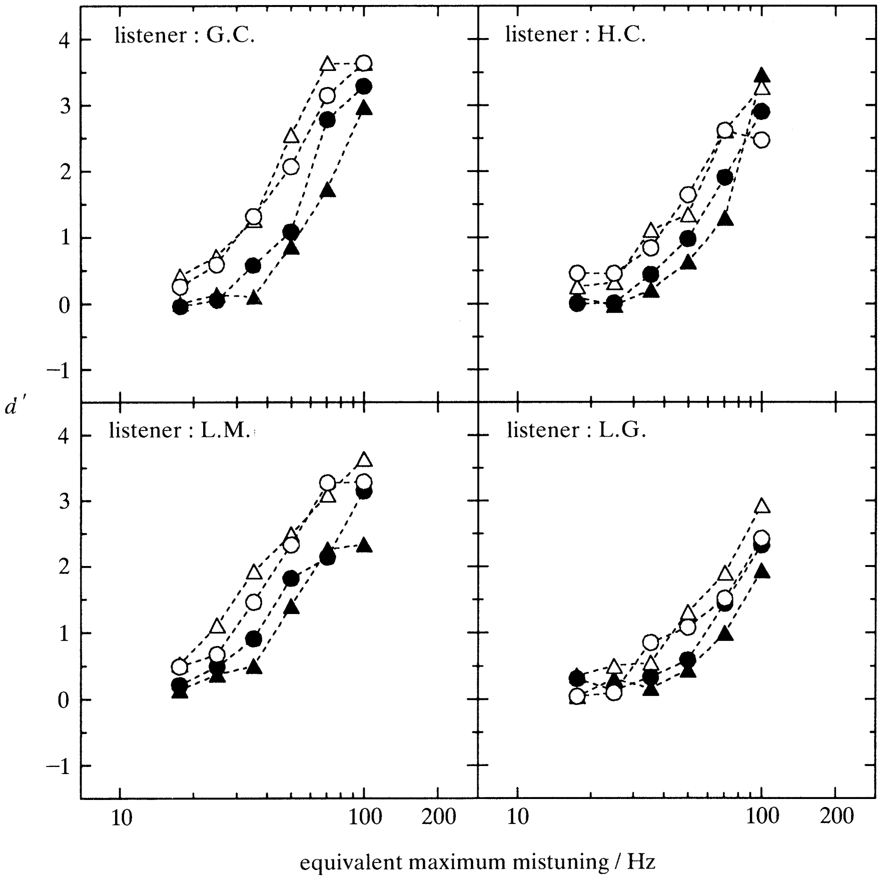

Figure 2. Sensitivity (d') as a function of modulator phase delay for each of four listeners. Data obtained from sinusoidal FM is shown by circles, that from triangular FM by triangles. Filled symbols are for the incoherence condition, and open symbols are for the mistuning condition.

paradigm with feedback was used. Stimuli consisted of three sinusoidal components (1500, 2000 and 2500 Hz), each modulated by a 2.5 Hz sinusoidal or triangular wave, with a zero-peak modulation depth of 2.5% of the carrier frequency. Signal duration was 400 ms, and all stimuli were presented at a level of 45 dB SPL per component in a background of pink noise (spectrum level 12.2 dB SPL at 2000 Hz).

The standard (non-signal) stimulus always consisted of the three sinusoidal components, in tune and modulated coherently. In the 'mistuning' conditions, sensitivity (d') was measured as a function of a static mistuning (± 18, 25, 35, 50, 71, 100 Hz) imposed on the 2000 Hz component in the signal interval. In the 'incoherence' conditions, d' was measured as a function of the modulator phase delay (0.11 to $1.0\,\pi$ radians) imposed on the 2000 Hz component. The modulator delays were chosen so that the maximum mistunings that they produced (at any time during the signal) were equivalent to the static mistunings imposed in the other condition (Carlyon 1991); accordingly, data in both conditions will be discussed and plotted in terms of 'equivalent maximum mistunings (EMMS)'. Plotting data in this way allows us to make a specific prediction: if listeners detect FM incoherence from the mistuning that it causes, then the psychometric functions in the mistuning and incoherence conditions should be parallel, as we would expect sensitivity to vary with EMM in the same way for both conditions. In all conditions, the playback rate, and hence the nominal F_0, was randomized by ± 5% from presentation to presentation, and the overall modulator starting phase was randomized from zero to π radians.

(*c*) *Results*

Sensitivity (d') is shown as a function of EMM for each of the four listeners in figure 2. One important comparison is in the incoherence condition, between performance with sinusoidal (filled circles) and with triangular (filled triangles) modulation. Performance for all four listeners is consistently better with sinusoidal modulation, indicating that they were indeed detecting mistuning, rather than FM incoherence per se (planned comparisons†, $F = 7.449$, $p < 0.05$). This conclusion is supported by the observation that for both types of modulation, the slopes of the psychometric

---

† The planned comparisons were performed after a two-way ANOVA (four 'modulation conditions' (combinations of modulator type and incoherence or mistuning) × six 'equivalent maximum mistunings (EMMS)'). The ANOVA revealed significant main effects of modulation condition ($F = 22.055$, $p < 0.001$), of EMM ($F = 64.042$, $p < 0.001$). The interaction between modulation condition and EMM was not significant.

functions in the incoherence (filled symbols) and mistuning (open symbols) conditions do not differ significantly (*t*-tests, c.f. Edwards 1973). Both of these results support the conclusions drawn in the Carlyon (1991) paper that, for harmonic stimuli, listeners detect FM incoherence from the mistuning that it causes. Taken together with the finding that listeners cannot detect FM incoherence imposed on a component of an inharmonic complex, this provides strong evidence against an across-frequency mechanism specific to the detection of FM incoherence.

3. MODULATION DETECTION INTERFERENCE

Elsewhere in this publication, Moore & Shailer describe a phenomenon, termed modulation detection interference (MDI), which reflects the across-frequency processing of complex sounds. In a typical MDI experiment, listeners are required to detect amplitude or frequency modulation of a sinusoidal carrier in the presence of an interfering sinusoid, whose frequency is such that it does not mask the signal. A common finding is that although the 'interferer' has no effect when it is unmodulated, a modulated interferer increases the signal's modulation detection threshold (Yost & Sheft 1989; Yost *et al.* 1989; Wilson *et al.* 1990; Moore *et al.* 1991). Thus, although the interferer has no effect on the signal in a detection experiment, it affects performance in a supra-threshold task such as the detection of modulation. In a recent paper, Wilson *et al.* (1990) reported that, for the detection of FM, threshold was affected not only by the imposition of FM on the interferer, but also by the amount of incoherence (modulator phase delay) between the signal and interferer modulation: coherent FM raised thresholds more than did incoherent FM. They suggested that, with coherent FM, listeners grouped the interferer with the signal, and that this impaired the detection of FM. With incoherent FM, the interferer would not be so strongly grouped with the signal, and so the FM detection threshold would be elevated less.

The dependence of MDI on FM coherence seems, at least at first sight, to contradict the conclusions drawn in the first part of this article: if listeners are not sensitive to FM coherence, how can it affect FM detection thresholds? Wilson *et al.*'s stimuli were inharmonic, with signal and interferer frequencies of 1900 Hz and 2500 Hz respectively, so it is unlikely that their results were due to differences in harmonicity between coherently and incoherently modulated sounds. However, there are two other possible reasons why they may have arrived at a different conclusion from that drawn here. Both arise from the observation that, whereas we presented our stimuli at a level of 45 dB SPL per component in a pink noise background, Wilson *et al.*'s sounds were presented at 65 dB SPL per component in quiet.

The first possibility was suggested by a pilot experiment in which I tried to discriminate between coherently and incoherently modulated versions of Wilson *et al.*'s stimili. I found that I could do so, but only by listening for a prominent combination tone (CT) in the incoherent case: when a 1400 Hz lowpass noise was added to mask the CTs, my performance dropped to chance. There are at least two ways in which CTs could affect the amount of MDI. First, in the incoherent condition, a CT with a frequency such as f_2-f_1 or $2f_1-f_2$ would be modulated over a much wider frequency range in the signal than in the non-signal interval, and listeners could detect this prominent-sounding CT, thereby lowering thresholds and hence the amount of MDI. Second, in the standard (non-signal) interval of a two-interval, forced-choice trial, the combination tone f_2-f_1 will be frequency modulated over the same range as the interferer. Moore *et al.* (1991) have reported that the existence of a second interfering tone increases the amount of MDI, and it is possible that the (modulated) CT raised modulation detection thresholds in a similar manner to an externally presented tone. As the stimuli in the standard interval are very similar in the coherent and incoherent conditions, this additional effect would be the same in the two conditions. Thus there are two possible effects of CTs; one which might reduce MDI in the incoherent condition, and a second which would tend to raise MDI in both conditions.

An alternative explanation for Wilson *et al.*'s finding arises from beating between the target and interfering tones in auditory filters that respond to both components. The beating is most likely to occur in filters with centre frequencies (CFs) between the target and interferer frequencies. In one of their experiments, Wilson *et al.* (1990) interspersed a narrowband noise between the target and interferer tones, and showed that the MDI caused by coherent FM persisted, and was therefore not mediated by within-channel beating. However, with incoherent FM, the within-channel beating will be stronger than in the coherent condition, and stronger than when the target is not modulated (i.e. in the non-signal interval). This is because during incoherent FM the component frequencies first move away from each other, causing the output of a filter tuned between them to decrease, and then move towards each other, causing the filter output to increase. Listeners might use this 'within-channel amplitude modulation' to detect the signal in the incoherent condition, thereby reducing their MDI. Thus, even though beating might not be necessary for the basic MDI phenomenon, it might account for the difference in MDI obtained with coherent and incoherent FM.

Experiment 2 investigated whether Wilson *et al.*'s finding of greater MDI for coherent than for incoherent FM was really due to an across-frequency grouping mechanism as they suggest, or whether it can be attributed to the detectability of CTs and within-channel beating.

(*b*) *Method*

The first set of conditions was a direct replication of Wilson *et al*'s experiment 3. Using an adaptive procedure (Levitt, 1971), the threshold FM depth necessary for listeners to discriminate between a 1900 Hz carrier that was either unmodulated, or sinusoidally frequency-modulated at a rate of 6 Hz, was measured in the presence of an interfering 2500 Hz tone. The interferer was presented in both intervals of each 2I, 2AFC trial, and was either unmodulated, modulated coherently with the signal (6 Hz, zero-peak FM depth fixed at 2.5% of carrier frequency), or modulated incoherently with the signal. Incoherent FM was produced by introducing a delay of π radians between the interferer and signal modulators. Both the interferer and the target component were presented at a level of 65 dB SPL in quiet. In a second set of conditions, CTS were masked by presenting the stimuli in a background of continuous lowpass noise, generated by passing a pink noise through two lowpass filters (Kemo VB25.03; attenuation 48 dB per octave each) in series. The noise had a spectrum level of 32.2 dB SPL at 1 kHz, and the filter cutoffs (3-dB-down) were set to 1400 Hz. In a third set of conditions, both CTS and within-channel interactions were masked by a 5-kHz-wide pink noise, whose spectrum level at 1 kHz was also 32.2 dB SPL.

(*c*) *Results*

The data for stimuli presented in quiet are shown in figure 3*a*. The existence of MDI is confirmed by the observation that, for four out of five listeners, thresholds are greater with a coherently modulated interferer (open bars), than when the interferer is unmodulated (filled bars). Also, the general pattern of results reported by Wilson *et al.* is confirmed, in that thresholds with incoherent FM are generally lower than with coherent FM (compare shaded with open bars). The contribution of CTS to this effect of FM coherence is shown in figure 3*b*, which shows that, when stimuli are presented in lowpass noise, the effect persists only for two out of the five listeners (G.C. and L.G.). Note that CTS do not seem to be essential for the basic MDI effect: thresholds obtained with an unmodulated interferer are still lower than in the presence of a coherently modulated interferer, even in the presence of lowpass noise. Finally, figure 3*c* shows that when within-channel interactions are also eliminated, by adding wideband pink noise instead of lowpass noise, all five listeners show essentially identical thresholds with coherent and incoherent interferers. For four listeners, thresholds with both of these modulated interferers are higher than in the unmodulated condition. The results of experiment 2 confirm that CTS and within-channel cues are not essential for the basic MDI phenomenon (Yost & Sheft 1989; Yost *et al.* 1989; Wilson *et al.* 1990; Moore *et al.* 1991), but that they are responsible for the dependence of MDI on FM coherence.

4. DISCRIMINATING BETWEEN TUNED AND MISTUNED HARMONICS IN THE PRESENCE OF A COMPETING SOUND

(*a*) *Motivation and general description*

The experiment described in the first part of this article added to the existing evidence that listeners can detect a mistuning imposed on one component of an otherwise harmonic complex tone (Moore *et al.* 1985*b*; Moore & Glasberg, 1986; Demany *et al.* 1991; Carlyon *et al.* 1992). The fact that we can 'hear out' a mistuned harmonic that is resolved from its neighbours shows that the auditory system can stimultaneously compare the frequencies of resolved harmonics, and can use this information to perceptually separate one component of a complex sound from the rest. Both Moore *et al.* (1985*b*) and Demany *et al.* (1991) have shown that listeners can detect mistunings of individual resolved components of as little as 1%; however, these measurements were made either in quiet (Moore *et al.* 1985*b*), or in the presence of a low-level pink-noise

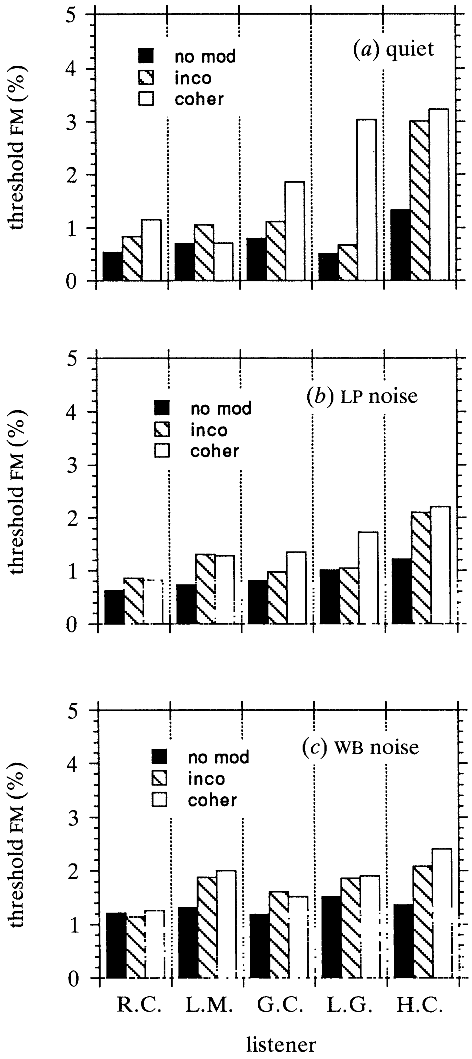

Figure 3. Each panel shows modulation detection thresholds, expressed as a percentage of the carrier frequency, for five listeners (spaced along the abscissa). For each listener, three thresholds are shown, corresponding to the modulation imposed on the interferer. Filled bars represent no modulation, open bars coherent modulation, and circles incoherent modulation. Modulation detection interference for the two conditions with modulated interferers is equal to the threshold elevation re. the condition with an unmodulated interferer. (*a*) Stimuli presented in quiet; (*b*) 1400 Hz lowpass noise background; (*c*) wideband pink noise background.

background (Demany *et al.* 1991). In real life, even when a component of a harmonic complex is resolved from its neighbours, there will sometimes exist a component of a different complex (e.g. a competing vowel sound) with a frequency very close to that of the target. One aim of our third experiment was to determine the extent to which detection of mistuning is affected by an interfering sound. We also investigated whether sensitivity to mistuning was affected by the existence of onset and offset asynchronies between the interfering sound and the harmonic complex.

(*b*) *Method*

Experiment 3 took advantage of the fact that one can measure sensitivity to mistuning by frequency modulating one component of a harmonic complex incoherently from the other components (cf. experiment 1). The harmonic complex consisted of the first seven harmonics of 500 Hz, presented at the same level and in the same background noise as the stimuli in experiment 1, and with a duration of 200 ms. In the standard interval of each 21, 2AFC trial, all components were frequency modulated coherently at a rate of 5 Hz; in the signal interval, the 'target component' (the fourth harmonic, frequency = 2000 Hz) was modulated incoherently (π radians modulator delay). The amount of mistuning is proportional to the FM depth imposed on all components, so the threshold FM depth was determined using an adaptive procedure (Levitt 1971).

In some conditions, there was an interfering sinusoid present in both intervals of each trial; its level was -9 dB, -6 dB, -3 dB, or 0 dB relative to that of the components of the harmonic complex. The 'interferer' was either turned on and off at the same time as the complex (condition 'SYNCH'; figure 4*a*), or was turned on 400 ms before its onset and turned off 100 ms after its offset (condition 'ASYNCH'; figure 4*b*). It should be obvious from figure 4*a*, *b*, that within-channel interactions (such as beating) will occur between the interferer and the target (2000 Hz) component, and that their patterning will depend on the phase of the target modulation. It is necessary to ensure that listeners do not identify the signal interval using such a within-channel cue. Therefore, the overall starting modulation phase of the stimuli was randomized from presentation to presentation: this is illustrated in figure 4*c*, which shows a different possible trial structure in the 'SYNCH' condition. Note that any particular pattern of interaction between target and interferer is equally likely to occur in the signal and standard intervals.

In an additional condition, the interfering sinusoid was replaced by a 200-Hz-wide noise, with an overall level 6 dB below that of the target component. It was gated on either synchronously or asynchronously with the target.

(*c*) *Results*

Figure 5 shows the threshold FM depth for the SYNCH (open triangles) and ASYNCH (open squares) conditions, as a function of the level of the interferer relative to that of the target component. In the SYNCH condition, thresholds are not substantially affected by interferers with levels re. the target of -9 or -6 dB, but rise steadily with further increases in interferer level. Perhaps the most striking finding is that, in the ASYNCH condition, thresholds are higher than in the SYNCH condition, and are affected by interferers with levels as low as -9 dB re. the target. This is true not only for an interfering sinusoid, but also for an interfering narrowband noise (compare the filled triangles and squares). The reasons for this additional effect of interferer asynchrony (over and above that caused by a synchronous interferer) will be discussed in § 5*b*.

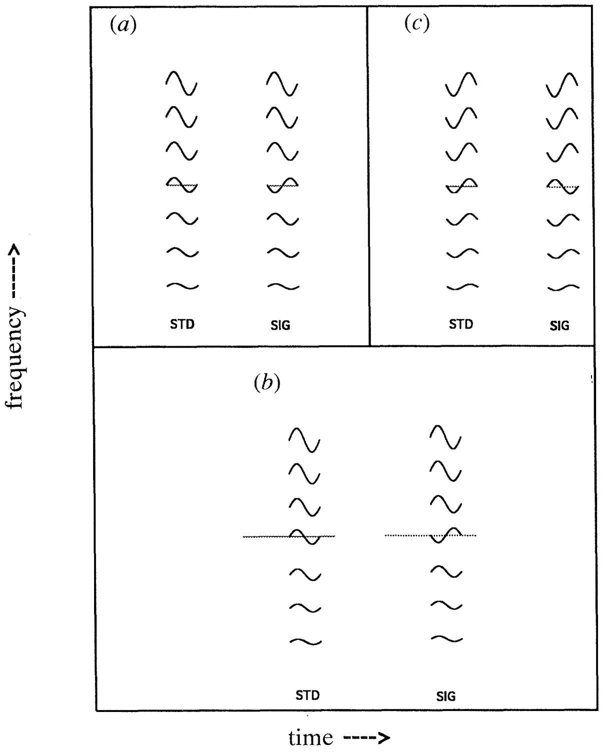

Figure 4. Panels (*a*) and (*c*) show schematic spectrograms of two possible trial structures in the SYNCH condition of experiment 3. Panel (*b*) shows one of the possible trial structures in the ASYNCH condition.

5. DISCUSSION

(*a*) *Why can't listeners discriminate coherent from incoherent FM?*

Experiments 1 and 2 supported the conclusion reached by Carlyon (1991) that there is not an across-frequency mechanism specific to the detection of FM incoherence. Given that FM incoherence seems such an attractive cue for segregation, the question arises as to why listeners cannot detect it in the absence of other cues. It seems likely that this is due to a number of reasons.

First, Summerfield (1991, see also this symposium) has pointed out that a mechanism specific to the detection of FM incoherence would be computationally

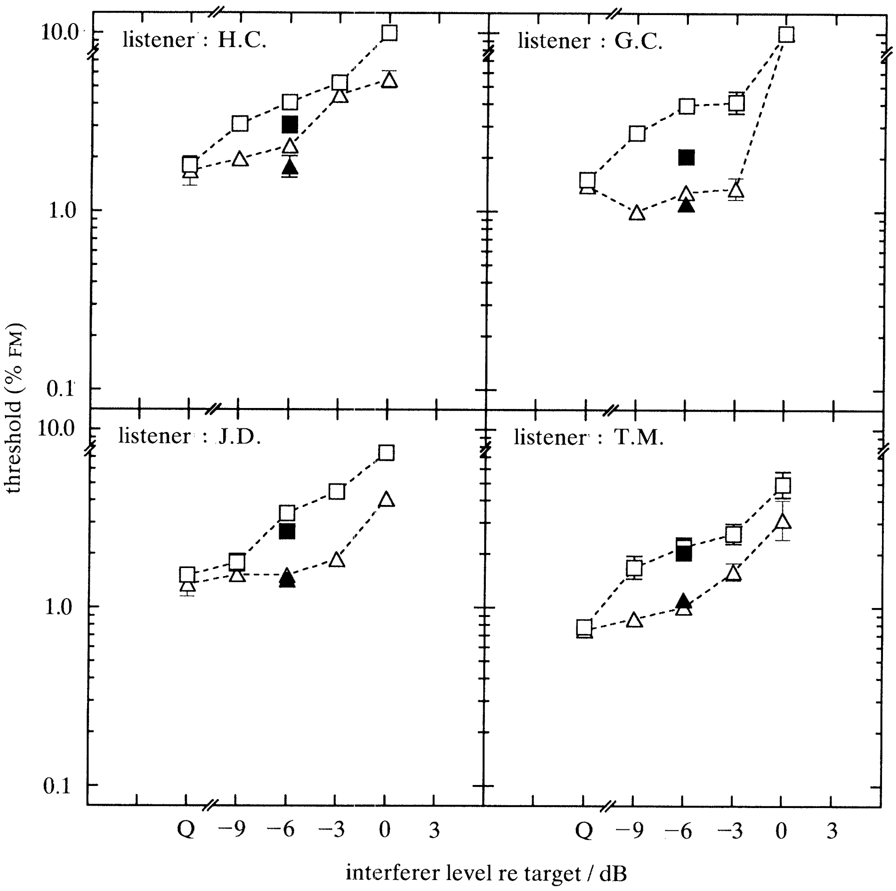

Figure 5. Modulation depth (as a percentage of carrier frequency) necessary for listeners to detect mistuning imposed on the fourth harmonic of a seven component complex ($F_0 = 500$ Hz). Each panel shows the data of one listener. Triangles, SYNCH condition; squares, ASYNCH condition. The open symbols are for a sinusoidal interferer, and the unconnected filled symbols are for a narrowband noise interferer.

expensive. He notes that most sounds that one would want to segregate on this basis are harmonic, and that as FM incoherence always leads to inharmonicity (for which there already exists a mechanism), it would be ecologically disadvantageous to develop a separate mechanism that provides no new information.

Second, it is worth noting recent evidence that coherent FM enhances grouping between a complex tone and an extra component, relative to the case with no FM. C. J. Darwin and V. Ciocca (personal communication) repeated an experiment by Moore *et al.* (1985), in which listeners were asked to adjust the pitch of an (unmodulated) harmonic complex to that of another complex which had had one component mistuned. They replicated Moore *et al.*'s finding for steady tones that, as the component was mistuned by up to about 3%, the pitch of the complex changed by increasing amounts in the same direction, whereas for larger mistunings the mistuned harmonic became segregated from the complex, and had a progressively smaller effect on its pitch. Darwin and Ciocca reported that when the whole stimulus was frequency modulated, the mistuned harmonic could contribute to the pitch of the complex at greater mistunings than was the case with no FM. They concluded that listeners have a tendency to group together components that are modulated. If modulated components are grouped together, then the effects of this grouping will be apparent only for coherent FM, as incoherent FM will lead to mistuning, which will in turn counteract any 'grouping by modulation'. Thus, by a combined sensitivity to mistuning and to modulation, listeners could fuse coherently modulated harmonic components, and separate incoherently modulated components. This strategy would be computationally more efficient than correlating the movements of peaks in the excitation pattern.

Finally, it is worth noting that a 'peak correlation' mechanism for detecting FM incoherence could produce the 'wrong answer' when the peaks reflect maxima in the spectral envelope (e.g. formants), rather than in the spectral fine structure (e.g. harmonics). Unlike harmonics, formants of the same speech sound often change frequency in opposite directions ('incoherently'), even when they come from the same source. Thus, in order to work effectively, a correlation mechanism would have to be able to take into account whether a given peak was due to a harmonic or to a formant.

In summary, it is proposed that the auditory system groups components by harmonicity, supplemented by the presence, but not by the coherence, of FM. Such a mechanism makes ecological sense: it is rare, but possible, for steady components from different sources to be harmonically related, but such a co-incidence is much less likely when the components are frequency modulated. Furthermore, unlike a 'peak-correlation' mechanism, the proposed processing scheme would be both computationally efficient and unlikely to erroneously separate groups of components that come from the same source.

(*b*) *The effect of interfering sounds on the detection of mistuning*

Experiment 3 showed that a sinusoid that starts 400 ms before, and ends 100 ms after, a harmonic complex impairs the detection of mistuning more than does a synchronous interferer. This finding seemed puzzling at first: one might expect listeners to use the leading portion of the interferer to 'subtract' it from the percept of the complex sound, thereby improving their detection of mistuning. At the very least, one might expect trained listeners such as ours to learn to ignore the leading and lagging portions of the tone. It is therefore worth considering reasons why they could not do this.

One possibility is that the leading portion of the tone induced adaptation in the frequency region of the target component, thereby reducing the level of its representation in the auditory periphery: it is known that frequency DLS increase at very low levels. However, figure 5 shows that the effect of onset asynchrony (ratio between ASYNCH and SYNCH thresholds) is as high at interferer levels of −9 dB as it is at −3 dB or 0 dB. If the leading portion of the interferer reduced the auditory-nerve response to the target component via adaptation, then we would expect the effect to be greatest at high interferer levels.

A second, more promising possibility is that the leading portion of the interferer captured the target component in the same perceptual 'stream' (see Bregman 1990), and that this impaired listeners' ability to compare the target frequency to that of the other components. A similar process might account for a recent finding reported by Green & Dai (1991), involving a different type of across-frequency comparison. They used a 'profile analysis' task, which requires listeners to detect an increment in the level of one component of a complex sound relative to that of the other components. They reported that threshold increased when the target component was turned on before, and off after, the rest of the complex compared to a condition with synchronous onsets and offsets. Further evidence for the central origin of the present effect comes from additional experiments in which the interfering sound was presented diotically, while the complex and target tones were presented monotically. These showed that mistuning detection was much worse than when all sounds were presented monotically. This may have been due to components with frequencies close to that of the interferer being pulled even more strongly into a separate perceptual stream.

It is notable that our listeners showed a synchrony effect even though the interferer and target 'sounded different' from each other when presented in isolation (one was a steady tone, the other frequency modulated), and that the effect persisted when this quality difference was increased by replacing the interferer with a narrowband noise. A possible interpretation is that a stimulus with a similar excitation pattern as the interferer is captured into its perceptual stream, irrespective of its other properties (tone versus noise, steady versus FM). If so, then in our experiment the auditory system acted in a fairly unsophisticated way when forming the continuous energy around 2000 Hz into a single perceptual stream. This outcome might be related to the results of Roberts & Moore (1990), who showed that adding pairs of tones with frequencies slightly above the first formant of a vowel increased the perceived first formant frequency (all sounds were gated on and off together). They reported that a similar effect could be obtained by replacing the tone pairs with a narrowband noise, despite its 'radically different temporal structure and timbre' from the vowel sound.

In conclusion, experiment 3 has shown that sensitivity to mistuning persists not only when the harmonics are resolved by the peripheral auditory system, but also when a synchronous narrowband sound is present whose spectrum overlaps that of the mistuned component. However, this 'resilience' to the interferer can be impaired by other mechanisms when the interferer starts before, and ends after, the harmonic complex. The exact nature of the additional mechanism is yet to be determined, but we know that it acts centrally and is fairly insensitive to temporal properties of the interfering sound (e.g. tone versus noise). Further investigations, particularly into the reasons why even highly-trained listeners cannot 'ignore' the leading and lagging portions of the interferer, are currently underway.

The work reported here was supported by a Royal Society University Research Fellowship to the author. I thank Valter Ciocca and Quentin Summerfield for useful comments on an earlier version of this article. In particular, Valter Ciocca provided useful insights into the interpretation of experiment 3.

REFERENCES

Bregman, A.S. 1990 *Auditory scene analysis*. Cambridge, Massachusetts: M.I.T. Press.

Bregman, A.S., Levitan, R. & Liao, C. 1990 Fusion of auditory components: effects of the frequency of amplitude modulation. *Percept. Psychophys.* **47**, 68–73.

Brokx, J.P.L. & Nooteboom, S.G. 1982 Intonation and the perceptual separation of simultaneous voices. *J. Phonet.* **10**, 23–36.

Carlyon, R.P. 1991 Discriminating between coherent and incoherent frequency modulation of complex tones. *J. acoust. Soc. Am.* **89**, 329–340.

Carlyon, R.P., Demany, L. & Semal, C. 1992 Detection of across-frequency differences in fundamental frequency. *J. acoust. Soc. Am.* (In the press.)

Carlyon, R.P. & Stubbs, R.J. 1989 Detecting single-cycle frequency modulation imposed on sinusoidal, harmonic, and inharmonic carriers. *J. acoust. Soc. Am.* **85**, 2563–2574.

Demany, L. & 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. & Semal, C. 1991 Harmonic and melodic octave templates. *J. acoust. Soc. Am.* **88**, 2126–2135.

Demany, L., Semal, C. & Carlyon, R.P. 1991 Perceptual limits of octave harmony and their origin. *J. acoust. Soc. Am.* **90**, 3019–3027.

Edwards, A.E. 1973 *Statistical methods.* New York: Holt, Rhinehardt, Winston.

Fletcher, H. 1940 Auditory patterns. *Rev. mod. Phys.* **12**, 47–65.

Green, D.M. 1988 *Profile analysis.* New York: Oxford University Press.

Green, D.M. & Dai, H. 1992 Temporal relations in profile comparisons. In *Auditory physiology and perception* (ed. Y. Cazals, L. Demany & K. Horner), pp. 471–478. Oxford: Pergamon Press.

Hall, J.W., Haggard, M.P. & Fernandes, M.A. 1984 Detection in noise by spectro-temporal pattern analysis. *J. acoust. Soc. Am.* **76**, 50–56.

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

Levitt, H. 1971 Transformed up-down methods in psychophysics. *J. acoust. Soc. Am.* **49**, 467–447.

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

Moore, B.C.J. & Glasberg, B.R. 1986 Thresholds for hearing mistuned partials as separate tones in harmonic complexes. *J. acoust. Soc. Am.* **80**, 479–483.

Moore, B.C.J., Glasberg, B.R., Gaunt, T. & Child, T. 1991 Across-channel masking of changes in modulation depth for amplitude- and frequency-modulated signals. *Q. J. exp. Psychol.* **43A**, 327–348.

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

Moore, B.C.J., Peters, R.W. & Glasberg, B.R. 1985*b* Thresholds for the detection of inharmonicity in complex tones. *J. acoust. Soc. Am.* **77**, 1861–1867.

Patterson, R.D. 1976 Auditory filter shapes derived with noise stimuli. *J. acoust. Soc. Am.* **59**, 640–654.

Roberts, B. & Moore, B.C.J. 1990 The influence of extraneous sounds on the perceptual estimation of first-formant frequency in vowels. *J. acoust. Soc. Am.* **88**, 2571–2583.

Scheffers, M.T.M. 1983 Sifting vowels: auditory pitch analysis and sound segregation. Ph.D. thesis, University of Groningen.

Summerfield, A.Q. 1991 Perception of concurrent vowels: effects of coherent frequency modulation. *Brit. J. Audiol.* (In the press.)

Summerfield, A.Q. & Assmann, P.F. 1991 Perception of concurrent vowels: Effects of pitch-pulse asynchrony and harmonic misalignement. *J. acoust. Soc. Am.* **89**, 1364–1377.

Wilson, A.S., Hall, J.W. & Grose, J.H. 1990 Detection of frequency modulation (FM) in the presence of a second FM tone. *J. acoust. Soc. Am.* **88**, 1333–1338.

Yost, W.A. & Sheft, S. 1989 Across-critical-band processing of amplitude-modulated tones. *J. acoust. Soc. Am.* **85**, 848–857.

Yost, W.A., Sheft, S. & Opie, J. 1989 Modulation interference in detection and discrimination of amplitude modulation. *J. acoust. Soc. Amer.* **86**, 2138–2147.

Zwicker, E., Flottorp, G. & Stevens, S.S. 1957 Critical bandwidths in loudness summation. *J. acoust. Soc. Am.* **29**, 548–557.

Discussion

S. McAdams (*Laboratoire de Psychologie Expérimentale, Université René Descartes, Paris, France*). The papers presented by Dr Summerfield and Dr Carlyon have presented rather convincing arguments against the direct involvement of coherent frequency modulation (CFM) in the perceptual segregation of concurrent sounds, the intimation being that FM does not account for any segregation over and above that which can be accounted for by the mistuning that such modulation would produce. If we accept this reasoning, however, there remains to be explained some puzzling experimental data. McAdams (1989) presented listeners with complex stimuli consisting of three vowels (/a/,/o/,/i/), each separated by five semitones (30%) which is well beyond the separation at which mistuning produces its maximum effect. Listeners were asked to judge the relative prominence of the vowels heard in the complex. Mean prominence judgments increased significantly when the vowels were modulated compared to when they were not. This finding suggests that something associated with FM made the vowels easier to hear in a complex background when modulated, in spite of the large static mistuning between them. The degree of increase in perceived prominence was the same when the vowels were: (i) modulated alone against a background of unmodulated vowels; (ii) modulated independently of the other vowels; or (iii) modulated coherently with the other vowels. Thus, whatever segregation was achieved by modulating a given vowel was no further enhanced by modulating the other concurrent vowels independently. Dr Carlyon's dichotomy between cues for grouping and cues for segregation claims that FM plays only a small role in grouping and none in segregation. According to this scheme then, the increased prominence would be due to the vowels' harmonics being perceptually grouped by FM. How, though, are we to explain the fact that grouping alone increases a vowel's perceived prominence in a complex mixture?

Reference

McAdams, S. 1989 Psychological constraints on form-bearing dimensions in music. *Contemp. Music Rev.* **4**, 181–198.

R. P. Carlyon. The finding Dr McAdams describes is of course consistent with those of Quentin Summerfield and myself, and represents some of the earliest and most convincing evidence that, although listeners are sensitive to the existence of FM, they are not sensitive to FM coherence between different concurrent sources.

It seems that, when a bunch of components are modulated by about the same amount, they tend to group together and become more prominent than when they are steady. In the sense that prominence can be considered equivalent to segregation, my arguments predict that FM can (and probably does) play a role in segregating a group of components from a background of other components. The only sense in which FM does not play a role in segregation is that its absence does not cause components to be segregated from each other in the way that, for example, the absence of harmonicity or of onset synchrony does. I think this distinction is an important one, and thank Dr McAdams for prompting me to make it more explicit.