Detection of combined frequency and amplitude modulation

Brian C. J. Moore
Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB, England

Aleksander Sek
Institute of Acoustics, Adam Mickiewicz University, 60-769 Poznan, Poland

(Received 6 April 1992; accepted for publication 6 August 1992)

This article is concerned with the detection of mixed modulation (MM), i.e., simultaneously occurring amplitude modulation (AM) and frequency modulation (FM). In experiment 1, an adaptive two-alternative forced-choice task was used to determine thresholds for detecting AM alone. Then, thresholds for detecting FM were determined for stimuli which had a fixed amount of AM in the signal interval only. The amount of AM was always less than the threshold for detecting AM alone. The FM thresholds depended significantly on the magnitude of the coexisting AM. For low modulation rates (4, 16, and 64 Hz), the FM thresholds did not depend significantly on the relative phase of modulation for the FM and AM. For a high modulation rate (256 Hz) strong effects of modulator phase were observed. These phase effects are as predicted by the model proposed by Hartmann and Hnath [Acustica 50, 297-312 (1982)], which assumes that detection of modulation at modulation frequencies higher than the critical modulation frequency is based on detection of the lower sideband in the modulated signal's spectrum. In the second experiment, psychometric functions were measured for the detection of AM alone and FM alone, using modulation rates of 4 and 16 Hz. Results showed that, for each type of modulation, d' is approximately a linear function of the square of the modulation index. Application of this finding to the results of experiment 1 suggested that, at low modulation rates, FM and AM are not detected by completely independent mechanisms. In the third experiment, psychometric functions were again measured for the detection of AM alone and FM alone, using a 10-Hz modulation rate. Detectability was then measured for combined AM and FM, with modulation depths selected so that each type of modulation would be equally detectable if presented alone. Significant effects of relative modulator phase were found when detectability was relatively high. These effects were not correctly predicted by either a single-band excitation-pattern model or a multiple-band excitation-pattern model. However, the detectability of the combined AM and FM was better than would be predicted if the two types of modulation were coded completely independently.

PACS numbers: 43.66.Fe, 43.66.Ba, 43.66.Lj

INTRODUCTION

The perception and detection of changes in the amplitude and frequency of sound has been extensively discussed in the psychoacoustic literature (Allanson and Newell, 1966; Coninx, 1978a; Feth, 1972; Hartmann and Hnath, 1982; Maiwald, 1967; Ozimek and Sek, 1987; Zwicker, 1952, 1956). These studies have led to two basic hypotheses on the perception of amplitude and/or frequency changes in an acoustic signal. One (Zwicker, 1956, 1962; Maiwald, 1967) postulates that a single mechanism is responsible for the perception of changes in amplitude and frequency. We will refer to this as the "single-mechanism" hypothesis. The other (Feth, 1972, Coninx, 1978a) postulates the existence of two independent mechanisms, one for amplitude changes and one for frequency changes. We will refer to this as the "independent-mechanisms" hypothesis.

The Zwicker-Maiwald model is essentially a place model based on the concept of the psychoacoustic excitation pattern (Zwicker, 1956, 1970). The excitation pattern of a sound can be defined as the output of the auditory filters as a function of center frequency, in response to that sound (Moore and Glasberg, 1983, 1987). The model assumes that changes in either amplitude or frequency are detected by monitoring the single point on the excitation pattern that changes most; this is equivalent to monitoring a single auditory filter. Detection of changes in amplitude or frequency occurs when the change in the excitation pattern exceeds a criterion amount, assumed by Zwicker (1956, 1970) to be approximately 1 dB.

Some more recent models assume that information can be combined over a certain region of the excitation pattern (equivalent to monitoring several auditory filters simultaneously) (Florentine and Buus, 1981; Moore and Glasberg,
whereas for the independent-mechanisms hypothesis rela-
dulators for AM and FM should play an important role, frequency modulation (FM). This type of modulation is frequency; both could be detected by virtue of the changes in the framework of these models, it is still possible to envisage a single detection mechanism for changes in amplitude and frequency; both could be detected by virtue of the changes in excitation level that they produce.

One approach to evaluating these models is to use signals with simultaneous amplitude modulation (AM) and frequency modulation (FM). This type of modulation is usually called mixed modulation (MM). Studies of this type have been carried out by Zwicker (1962), Allanson and Newell (1966), Feth (1972), Coninx (1978a,b), Hartmann and Hnath (1982) and Ozimek and Sek (1987). According to the single-mechanism hypothesis, one would expect that the perception of amplitude changes would depend on coexisting frequency changes and vice versa. According to the independent-mechanisms hypothesis, one would expect that perception of amplitude changes would be independent of coexisting frequency changes and vice versa. Also, for the single-mechanism hypothesis the relative phase of the modulators for AM and FM should play an important role, whereas for the independent-mechanisms hypothesis relative phase should be unimportant.

Zwicker (1962) compared the sensations produced by AM, FM, and MM of an octave-wide noise band. The upper side of its excitation pattern was masked by a second, fixed band of noise. The modulation rate was either 3 or 10 Hz. He found that the subjective modulation "strength" produced by FM could be either increased or decreased by the addition of AM, depending on the relative modulation phase of the FM and AM. He concluded that, for noise bands, the sensations produced by amplitude change and by frequency change are based on the same mechanism.

The perception of MM signals with clearly audible modulation depths was also investigated by Coninx (1978b). In addition, he measured thresholds for detecting combined differences in frequency and amplitude (Coninx, 1978a). In some cases, the modulated sinusoids were presented with a noise band intended to mask the upper side of the excitation pattern. In most cases, he found that the results were not affected by the relative phase of the modulators for AM and FM, which he interpreted as indicating essentially independent mechanisms for the detection and perception of FM and AM. Where phase effects were found, he interpreted these in terms of the influence of amplitude on pitch or frequency on loudness (Czajkowska et al., 1988).

Hartmann and Hnath (1982) extended the model advanced by Goldstein (1967) to account for differences in modulation perception at low and high modulation frequencies. For a sinusoidal carrier and low modulation frequencies, the percept is of a pure tone that is changing in pitch or loudness. For high modulation frequencies, for which the spectrum of the modulated signal covers a frequency range greater than one critical band, the percept is of a steady sound that may have an impure tone quality. Their results suggested that modulation detection at high modulation rates is based exclusively on the lower sideband of the spectrum, since the higher sideband is completely masked. This applied to AM, FM, and MM signals, consistent with a single-mechanism hypothesis for high modulation frequencies.

Ozimek and Sek (1987) also investigated the detectability of MM, using both low and high modulating frequencies. For high modulating frequencies, they found, like Hartmann and Hnath (1982), that detection of modulation is based exclusively on the lower sideband of the modulated signal. They also found that, at low modulation frequencies, amounts of FM and AM that were separately "sub-threshold" could be detected. However, it is unclear whether this reflects the ability of subjects to combine information from independent sources, or whether it reflects a summation of sensations, as assumed in the single-mechanism hypothesis. Ozimek and Sek only used in-phase modulation for the MM signal, so it is not known whether modulator phase affects the results, as would be expected from the single-mechanism hypothesis.

This overview of the literature indicates that, at least for low modulation rates, there is no clear consensus as to whether the single-mechanism hypothesis or the independent-mechanisms hypothesis is correct. This article reports an attempt to evaluate these mechanisms by studying the detection of (MM). In the first experiment, thresholds were initially determined for detecting AM alone and FM alone, using a two-alternative forced-choice task. Then, thresholds for detecting FM were determined for stimuli which had a fixed amount of AM in the signal interval only. The amount of AM was always less than the threshold for detecting AM alone. The relative phase of modulation for the AM and FM was systematically varied.

I. TEMPORAL AND SPECTRAL STRUCTURE OF A MIXED MODULATION SIGNAL

In this section, we give a brief description of the temporal and spectral structure of the MM stimuli used in our experiments. Following Ozimek and Sek (1987), the instantaneous amplitude of a MM signal can be written

\[
A(t) = A_0 \left[ 1 + m \cos(\omega_m t + \Phi) \right]
\times \cos(\omega_0 t + \beta \sin(\omega_m t + \Theta)),
\]

where \(t\) is time, \(A_0\) is the envelope amplitude of the unmodulated signal, \(m\) and \(\beta\) are the modulation indices for AM and FM, respectively, \(\omega_m\) is the modulation frequency, \(\omega_0\) is the carrier frequency, and \(\Phi\) and \(\Theta\) are terms representing the phase of the modulation. The phase difference between the modulating signals, \(\Delta\phi\), is \(\Phi - \Theta\). When \(\Delta\phi = 0\), a maximum in frequency coincides with a maximum in amplitude. In our experiments, we used values for \(\Delta\phi\) of 0, \(\pi/2\), \(\pi\), and \(3\pi/2\). Following the arguments presented by Ozimek and Sek (1987), it may be shown that the spectrum of a MM signal consists primarily of three components of which the middle one corresponds to the carrier, while the two sidebands are the result of the modulation process. The relative amplitudes and phases of the sidebands depend on the value of \(\Delta\phi\).

Figure 1(a)–(d) illustrates the temporal and spectral structure of MM signals for \(\Delta\phi = 0\), \(\pi/2\), \(\pi\), and \(3\pi/2\), re-
Fig. 1. The time pattern (top) and spectral structure (bottom) of mixed modulation signals. Each panel shows a different relative phase of the modulators for AM and FM: (a) $\Delta \phi = 0$; (b) $\Delta \phi = \pi/2$; (c) $\Delta \phi = \pi$; (d) $\Delta \phi = 3\pi/2$.

spectively. Each panel shows: an expression for the instantaneous amplitude of the MM signal in terms of its three largest Fourier components; a portion of the waveform for $m = 0.5$ and $\beta = 20$ (values of these parameters were selected so that amplitude and frequency changes could easily be seen); the spectrum of the MM signal; and expressions for the amplitude values of the sidebands. When $\Delta \phi = 0$ [Fig. 1(a)], the amplitudes of the sidebands are proportional to the sum (upper sideband) and difference (lower sideband) of the modulation indices $m$ and $\beta$. When $\Delta \phi = \pi/2$ [Fig. 1(b)], the sidebands have equal amplitudes proportional to the vector sum of the two modulation indices. When $\Delta \phi = \pi$
of the sidebands are proportional to the difference (upper sideband) and sum (lower sideband) of the modulation indices $m$ and $\beta$. When $\Delta \phi = 3\pi/2$ [Fig. 1(d)], the amplitude spectrum is the same as when $\Delta \phi = \pi/2$; the sidebands have equal amplitudes proportional to the vector sum of the two modulation indices. However, the phase spectrum is different for $\Delta \phi = \pi/2$ and $\Delta \phi = 3\pi/2$.

II. EXPERIMENT 1: THRESHOLDS FOR DETECTING MIXED MODULATION

A. Method

1. Stimuli

The carrier signal was always a sinusoid with frequency $f_0 = \omega_0/2\pi = 1000$ Hz and level 70 dB SPL. The modulator was a sinusoid with frequency $f_m = \omega_m/2\pi = 4, 16, 64,$ or 256 Hz. The values for $f_m$ were selected so as to cover three different perceptual ranges. For $f_m = 4$ and 16 Hz, the changes in amplitude and frequency can be heard as such. For $f_m = 64$ Hz, the fluctuations are too fast to follow, and the sound has a certain "roughness" (Terhardt, 1974; Kemp, 1982). For $f_m = 256$ Hz, the sound quality becomes smoother, and individual sidebands may be perceived (Hartmann and Hnath, 1982; Ozimek and Sek, 1987).

The signals were digitally generated using a Masscomp 5400 computer system via a 16-bit digital-to-analog converter (DAC, model DA04H) at a sampling frequency of 10 kHz. The output of the DAC was low-pass filtered (Fern EF16, 100 dB/oct) with a cutoff frequency of 4 kHz, passed through a manual attenuator and delivered to one earpiece of a Sennheiser HD 414 headset.

On each trial, two successive stimuli were presented, one modulated and the other unmodulated. The order of the two stimuli in each pair was random. Each stimulus had an overall duration of 1000 ms, including raised-cosine rise/fall times of 50 ms. The time interval between the stimuli was 500 ms.

Initially, thresholds for detecting AM alone and FM alone were determined for each modulation frequency. The threshold modulation index for AM will be denoted by $m_{th}$ and that for FM by $\beta_{th}$. Next, thresholds were measured for detecting FM, when that FM was combined with "sub-threshold" amounts of AM at the same modulation frequency. In this stage of the experiment, the stimulus in the signal interval was both amplitude and frequency modulated, while the other stimulus was unmodulated. Within a block of trials, the amount of AM ($m$) was held constant, and the amount of FM ($\beta$) was varied to determine threshold. Values of $m$ used were 0.2$m_{th}$, 0.4$m_{th}$, 0.6$m_{th}$, 0.8$m_{th}$, and $m_{th}$. Four values of $\Delta \phi$, the phase shift between the modulators for AM and FM, were used, namely 0, $\pi/2$, $\pi$, and $3\pi/2$.

2. Procedure

An adaptive two-alternative forced-choice (2AFC) procedure was used. The modulation index (either $m$ or $\beta$) was increased after one incorrect response and decreased after three successive correct responses. This procedure tracks the point on the psychometric function corresponding to 79.4% correct (Levitt, 1971). The modulation index was changed by a factor of 1.5 until four reversals had occurred, and by a factor of 1.26 for the rest of the run. For each threshold measurement, 12 turnpoints were obtained; the threshold was calculated as the geometric mean value of the modulation index at the last eight turnpoints. Thresholds presented are the average of at least five (usually more than five) single threshold determinations. When the standard deviation of the mean value was more than 15% of the mean, at least one additional threshold estimate was obtained, and all estimates were averaged.

Subjects were tested in a double-walled sound-attenuating chamber. Their task was to indicate (by pushing an appropriate button) whether the first or the second stimulus in a pair was modulated in amplitude or frequency. Correct-answer feedback was provided by lights on the response box.

3. Subjects

Three subjects with normal hearing at all audiometric frequencies were used. One was author AS. The other two subjects were paid for their services. Two subjects (AS and RO) were experienced in similar tasks, while one, KR, had no previous experience. All subjects were trained until their performance appeared to be stable.

B. Results

Thresholds for detecting AM alone and FM alone are shown in Fig. 2. The results are very similar across subjects and are consistent with those obtained by Zwicker (1952), Goldstein (1967), Schorer (1986), and Ozimek and Sek (1987). When AM and FM sounds have equal modulation
indices \( m = \beta \), and when the indices are small, the AM and FM sounds have the same amplitude spectrum but different phase spectra. For the two higher modulation rates, the phase spectrum evidently plays no role, since, at threshold, the modulation index is the same for AM and FM. However, for the two lowest modulation rates, the thresholds are clearly lower for AM than for FM, indicating a monaural phase effect. The critical modulation frequency is between 16 and 64 Hz, consistent with the data of Schorer (1986), who found a critical modulation frequency slightly less than 64 Hz for a 1-kHz carrier at 50 dB SPL.

The thresholds for detecting FM in the presence of fixed “sub-threshold” amounts of AM are shown in Fig. 3 for modulation frequencies up to 64 Hz. The abscissa shows the fixed amount of AM as a proportion of the “threshold” value (as shown in Fig. 2) for each subject, and the ordinate shows the amount of FM required for threshold with the MM stimulus, plotted as a percentage of the threshold for FM alone. The parameter is the phase shift between the AM and FM, \( \Delta \phi \). Each column shows results for one modulation frequency, while each row shows results for one subject.

The results for the cases where the fixed amount of AM was equal to the threshold for AM alone (100% on the abscissa) need to be treated with caution. In theory, the AM alone should have been sufficient to give the 79.4% correct responses needed for “threshold.” However, because the value of \( \beta \) was changed by a certain factor in the adaptive procedure, and was never allowed to become zero or negative, a positive value of \( \beta \) was always obtained for “threshold” in the MM case. The “thresholds” in these cases were not properly determined by the adaptive procedure and were highly variable; that is why the points in the figure corresponding to these thresholds are not connected by the lines through the other points. This problem may also have had some effect on the thresholds estimated when the amount of AM was a high proportion of its threshold value.

For fixed amounts of AM less than the threshold for AM alone (points joined by lines in Fig. 3), the amount of FM required for threshold decreased as the fixed amount of AM increased. This effect was rather large at the lowest modulation frequency (4 Hz, left column), and decreased somewhat as the modulation frequency increased to 16 Hz (middle column) and 64 Hz (right column). However, there was no clear effect of the relative modulator phase, \( \Delta \phi \).

The thresholds for fixed amounts of AM less than the threshold for AM alone were subjected to an analysis of variance (ANOVA) with factors subjects, modulation rate, AM depth, and relative modulator phase. The variance associated with the four-way interaction was used as an estimate of the residual variance. The main effects of subjects, modulation rate, and AM depth were all highly significant \( (p < 0.001) \). Thresholds were lower for the lower modulation rates and the higher values of AM depth. However, the effect of relative modulator phase was not significant; \( F(3,36) = 1.23, p = 0.308 \).

Since the results were very similar across subjects and across the values of \( \Delta \phi \), they were averaged across subjects and phase values to give a clearer indication of the effects of the fixed amount of AM and of modulation rate. The results are shown in Fig. 4. The dashed line shows the predictions of a model that will be described later. These averaged data show an orderly decrease in the amount of FM required for

![FIG. 3. Thresholds for detecting FM in the presence of a fixed amount of AM, for modulation frequencies of 4, 16, and 64 Hz. Thresholds are expressed as a percentage of the threshold measured for the detection FM alone. The amount of AM is expressed as a percentage of the threshold for detecting AM alone. Each symbol shows results for a different value of \( \Delta \phi \), the phase shift between the modulators for AM and FM.](image_url)

![FIG. 4. As Fig. 3, but showing results averaged across subjects and modulator phases. Each symbol shows results for a different modulation frequency. The dashed line shows predictions of a model assuming independent mechanisms for the detection of AM and FM.](image_url)
threshold as the fixed amount of AM is increased. They also confirm that the decrease tends to be smaller as the modulation rate increases.

Figure 5 shows results for each subject using a modulation frequency of 256 Hz. Here, there were very large effects of relative modulator phase, $\Delta \phi$. For $\Delta \phi = 0$, i.e., when the maxima in amplitude and frequency were coincident, an increase in AM depth ($m$) caused an increase in the FM index ($\beta$) required for threshold. In other words, the coexisting AM made the FM harder to detect. An opposite effect was observed for $\Delta \phi = \pi$; an increase in $m$ caused a significant decrease the value of $\beta$ required for threshold. For $\Delta \phi = \pi/2$ or $3\pi/2$, the values of $\beta$ required for threshold decreased slightly with increasing $m$.

At this modulation rate, the sidebands in the spectrum would probably be resolved in the auditory system. Thus a description of the modulated signal in terms of its temporal structure (i.e., as changes in frequency and amplitude over time) is not appropriate in this case. The results shown in Fig. 5 can be predicted on the basis of the spectral structure of the modulated signal and the frequency selectivity of the auditory system. These predictions are presented in Sec. VI B.

III. EXPERIMENT 2

In this experiment, we determined psychometric functions for the detection of AM alone and FM alone. The basic aim was to establish how the detectability index $d'$ varied with modulation depth. The results are used later in this paper to predict thresholds for the MM data presented above on the basis of the single-mechanism hypothesis and the independent-mechanisms hypothesis.

A. Method

The stimuli had the same frequency, level and time pattern as in experiment 1. Modulation frequencies of 4 and 16 Hz were used. The subjects were the same as for experiment 1. Unfortunately, subject AS was not able to complete testing at 16 Hz. Initially, we tried using a 2AFC task with the modulation index fixed for a block of 55 trials, treating the first five trials as practice. However, we found that performance at low modulation depths was often erratic, and worse than would be expected from the thresholds measured using the adaptive procedure of experiment 1. It appeared that subjects had difficulty in "knowing what to listen for" when the modulation depth was fixed at a low value (for a similar effect in simple signal detection, see Taylor and Forbes, 1983).

To overcome this problem, we modified the procedure so that five different modulation depths were used in each block of 55 trials. The highest modulation depth was chosen to be rather easily detectable (typically giving 93%-99% correct responses), and the first five trials within a block were practice trials at this modulation depth, to help the subject to "home in" on the appropriate detection cues. The remaining 50 trials were test trials. In the first five of these, the modulation depth started at the highest value and decreased progressively to the lowest. This was then repeated for the next five, and so on. Thus the subject was presented with an "easy" stimulus every five trials, as a reminder of what to listen for. The five values of modulation depth were chosen on the basis of pilot data so as to give values of $d'$ between about 0.25 and 2.8 (values of % correct between about 57 and 98). Twenty blocks of trials were run for each modulation frequency and type of modulation (AM or FM), so each point on each psychometric function is based on 200 judgments.

B. Results

The percent correct values were converted to values of $d'$ (Green and Swets, 1974; Hacker and Ratcliff, 1979). In a few cases, three or fewer errors were made in the 200 trials for a given condition. It is not possible to make an accurate estimate of $d'$ in such cases, so they were excluded from further analysis. Hence, the data presented are based only on stimuli for which the percent correct was 98% or less. Analysis of the results suggested that $d'$ was approximately a linear function of the modulation index squared, for both AM and FM. Hence, values of $d'$ are plotted as a function of $1000m^2$ or $\beta^2$ in Figs. 6 and 7. Left-hand panels show results for AM detection and right-hand panels show results for FM detection. The diagonal line in each panel shows the best-fitting straight line, constrained to pass through the origin. As can be seen, these lines generally fit the data rather well; all deviations from the fitted lines are within the limits of experimental error. These results are consistent with those reported by Sheft and Yost (1989) for the detection of AM using a sinusoidal carrier, and by Irwin (1989) for the detection of 10-Hz AM of Gaussian noise; their data also suggest that $d'$ is a linear function of $m^2$.

IV. EVALUATION OF THE INDEPENDENT-MECHANISMS HYPOTHESIS

In this section, we consider predictions based on the independent-mechanisms hypothesis, without assuming any specific detection mechanisms for AM or FM. Let $\beta_0$, corre-
Consider a situation like that used in experiment 1, where the AM modulation index is fixed at some proportion, \( P(\leq 1) \), of \( m_{\text{th}} \), and the value of \( \beta \) is adjusted to give 79.4% correct. The value of \( d' \) that would result from the AM alone is (from assumption 1),

\[
d'_{\text{AM}} = d' \beta^2.
\]

If assumption (2) is correct, then, at threshold,

\[
d'_{\beta m} = d' \gamma = \sqrt{d'_{\beta}^2 + d'_{m}^2} = \sqrt{d'_{\beta}^2 + (d' \gamma P^2)^2}.
\]

Squaring, and rearranging terms gives

\[
d'_{\beta}^2 = d' \gamma^2 (1 - P^4),
\]

and

\[
d'_{m} = d' \gamma \sqrt{1 - P^4}.
\]

From assumption 1, the value of \( \beta \) corresponding to \( d'_{\beta} \) is

\[
\beta_{\text{th}} \sqrt{(d'_{\beta}/d' \gamma)}.
\]

Hence, the value of \( \beta \) necessary for 79.4% correct when combined with AM with depth \( Pm_{\text{th}} \) is

\[
\beta_{\text{th}} (1 - P^4)^{0.25}.
\]

The dashed line in Fig. 4, presented earlier, shows the threshold values predicted from Eq. (5). The symbols and solid lines show results averaged for the three subjects with \( P = 0.0, 0.2, 0.4, 0.6, 0.8 \), and 1.0, for modulation frequencies of 4, 16, and 64 Hz. It can be seen that the obtained values fall somewhat below the predicted values for \( P = 0.4, 0.6 \), and 0.8, especially when \( \omega_{m/2\pi} = 4 \) Hz. This indicates that assumption (2) is not correct. Rather, performance when AM and FM are combined is better than predicted, supporting assumption (3) above; AM and FM seem to be coded partly on a common dimension.
V. EXPERIMENT 3: PSYCHOMETRIC FUNCTIONS FOR THE DETECTION OF COMBINED AM AND FM

The analysis presented in the preceding section suggests that the detectability of combined AM and FM, as measured in experiment 1, is better than would be expected if AM and FM were coded independently. However, the adaptive threshold procedure used in experiment 1 was not ideal for testing the independent-mechanisms hypothesis, since the procedure probably did not "track" properly at the higher AM depths. In this experiment, we started by measuring psychometric functions for AM and FM alone, as in experiment 2. Then we selected pairs of values of AM and FM that would be equally detectable if presented alone. Psychometric functions were measured for the combined AM and FM, using the same four relative modulator phases as in experiment 1.

A. Method

Stimuli were similar to those used in experiments 1 and 2. However, to make data collection more rapid, the stimulus steady state duration was decreased to 500 ms and the interstimulus interval was decreased to 300 ms. The modulation rate was 10 Hz, chosen to be high enough that several modulation cycles would occur in each stimulus, but low enough that the modulation could be easily followed. Initially, psychometric functions were determined for AM alone and FM alone, using the same method as for experiment 2. Modulation indices were chosen on the basis of pilot experiments so as to yield a series of \( \Delta' \) values up to about 2.5. In most cases, five-point functions were determined. For one subject, a second five point series overlapping the first was obtained for AM detection only.

As in experiment 2, the data showed a linear relationship between \( \Delta' \) and the square of the modulation index. The data for AM alone and FM alone were fitted with functions of the form:

\[
d' = S_{AM} m^2 \tag{6}
\]

and

\[
d' = S_{FM} \beta^2 \tag{7}
\]

The best-fitting values of the slopes, \( S_{AM} \) and \( S_{FM} \), were used to derive pairs of values of \( m \) and \( \beta \) that would be equally detectable if presented alone. A series of such pairs of values was then used with each subject to determine the detectability of combined AM and FM. The psychometric functions for combined AM and FM were obtained using the same method as described earlier, where "easy" and "difficult" stimuli were intermingled within a block of trials. Four values of \( \Delta \phi \), the phase shift between the modulators for AM and FM, were used, namely 0, \( \pi/2 \), \( \pi \), and 3\( \pi/2 \).

Three young subjects were used. Two were undergraduates gathering the data for a research project. The other was paid for his services. All subjects had thresholds close to 0 dB HL at all audiometric frequencies. Subjects were trained for at least 4 h each, after which their performance appeared to be stable. Each point on each psychometric function is based on 200 trials.

B. Results

Psychometric functions for the detection of AM alone and FM alone are shown in the left- and right-hand panels of Fig. 8, respectively. The lines are fitted to the data assuming that \( d' \) is proportional to the square of the modulation index. There are no systematic deviations from the fitted values. Subject DH tended to be more sensitive than the other subjects, particularly for the detection of FM. This may reflect the fact that he had some musical training, whereas the other subjects did not.

Figure 9 shows psychometric functions for combined AM and FM. The abscissa shows the value of \( d' \) for the AM or FM alone, as given by the fitted lines in Fig. 8. The dashed line shows where the data should fall if the FM and AM were coded independently, and the information from them were combined optimally [Eq. (2)]. It is clear that the data fall consistently above the dashed line, suggesting that AM and FM are not coded completely independently. This is consistent with the conclusion reached earlier, on the basis of results from experiments 1 and 2.

In contrast with the results of experiment 1, the data do show evidence of an effect of modulator phase. This effect is clearest for the higher values of \( d' \), and is small or absent for
low values of $d'$. This probably accounts for our failure to find a phase effect in experiment 1, since the adaptive-procedure used in that experiment tracked a relatively low value of $d' (1.16)$. For all three subjects, the values of $d'$ are greatest for relative modulator phases, $\Delta \phi$, of 0 and $3\pi/2$. The values of $d'$ tend to be lowest for $\phi = \pi$, although for subject BB, $d'$ values were similar for $\phi = \pi$ and $\pi/2$.

To assess the statistical significance of these effects, a separate ANOVA was conducted for each subject. To obtain an estimate of the underlying variance of the data, the 200 trials used to estimate each $d'$ value were split into two lots of 100 trials, and $d'$ was calculated separately for each one (split-halves analysis). The ANOVA was then conducted with factors modulation depth and relative modulator phase, using the pairs of values as repeated measures. For all three subjects, the main effects of modulation depth and phase were significant ($p < 0.003$) but the interaction of modulation depth and phase was not significant at the 0.05 level. Post-hoc analyses revealed that, for each subject, the $d'$ values for $\Delta \phi = 0$ and $3\pi/2$ were not significantly different, but both were significantly greater than the $d'$ values for $\Delta \phi = \pi$ ($p < 0.02$ for BB and DH and $p < 0.05$ for HM). For subject BB the $d'$ values for $\Delta \phi = 0$ and $3\pi/2$ were significantly greater than for $\Delta \phi = \pi$ ($p < 0.02$), but for subjects DH and HM, the $d'$ values for $\Delta \phi = \pi/2$ were not significantly different from those for the other phase values.

In summary, the detectability of combined AM and FM was greater than would be predicted from the optimal combination of independent sources of information. There were significant effects of modulator phase, with modulator phases 0 and $3\pi/2$ giving the highest detectability, and $\pi$ tending to give the lowest.

VI. EVALUATION OF MODELS OF MODULATION DETECTION

In this section, we discuss the implications of the results for models of modulation detection. The discussion is separated into two parts, one concerned with modulation at relatively slow rates, where the changes of the stimulus over time can be followed by the auditory system, and the other concerned with modulation at high rates, where the time changes cannot be followed as such, and performance is based on the detection of spectral changes.

A. Detection of modulation at low rates

1. The Zwicker single-mechanism model

The model proposed by Zwicker (1952, 1956, 1970) assumes that detection of AM or FM depends on monitoring the single place on the excitation pattern that changes most. However, this place can be different for AM and FM. Generally, the largest change in excitation level for AM occurs on the high-frequency side of the pattern, whereas the largest change for FM occurs on the low-frequency side (Moore and Glasberg, 1986; Zwicker, 1956). It is possible that, for a MM signal, detection still depends on monitoring the single place on the excitation pattern that changes most. However, the location of that place would be expected to shift depending on the value of the phase difference between the AM and FM modulators, $\Delta \phi$.

If $\Delta \phi = 0$, then for a MM signal, the changes produced by AM and FM on the low-frequency side of the excitation pattern would tend to cancel (a rise in frequency producing a decrease in excitation level, and a rise in amplitude producing an increase), whereas the changes on the high-frequency side would add, producing a greater net change. Thus, in this case, detection should be based on monitoring the high-frequency side of the pattern. In contrast, if $\Delta \phi = \pi$, the changes produced by AM and FM on the low-frequency side of the excitation pattern would tend to add, whereas the changes on the high-frequency side would cancel. Thus, in this case, detection should be based on monitoring the low-frequency side of the pattern. For both $\Delta \phi = 0$ and $\Delta \phi = \pi$, the maximum change in excitation level produced by the MM signal is simply the sum of the changes produced by AM and FM individually on the appropriate side of the excitation pattern. For example, if each produced a change in relative amplitude of 0.06 (about 0.5 dB) at a given point, the overall change would be 0.12 (about 1 dB).
Consider now the situation for \( \Delta \phi = \pi/2 \) and \( 3\pi/2 \). Here, the AM and FM are orthogonal, so the maximum change in excitation level at any place is always less than the simple sum of the changes produced by AM alone and FM alone. Essentially, the powers of the modulators add, rather than their amplitudes. For example, if the AM alone and the FM alone each produced an amplitude change of 0.06 at a given place, the overall change would be 0.085 (about 0.71 dB). If performance were based on the detection of changes at a single place on the excitation pattern, performance should have been worse when \( \Delta \phi = \pi/2 \) or \( 3\pi/2 \) than when \( \Delta \phi = 0 \) or \( \pi \). The results of experiment 3 are not consistent with this prediction. Performance tended to be better for \( \Delta \phi = 3\pi/2 \) than for \( \Delta \phi = \pi \). Thus the data are not consistent with the assumption that performance is based on changes in excitation level at a single point on the excitation pattern.

2. A multiband excitation-pattern model

We evaluate here a multiband excitation-pattern model for the detection of modulation. Like Zwicker’s model, the model assumes that the detection of AM, FM, and MM are all based on a single form of information, namely changes in excitation level. However, this model assumes that information from different parts of the excitation pattern can be combined. The model is similar in its general form to the model for intensity discrimination proposed by Florentine and Buus (1981). It is based on the following assumptions.

1. AM, FM, and MM are all detected by virtue of the changes in excitation level that they produce in the peripheral auditory system.

2. If the change in excitation level in the \( i \)th critical band (or auditory filter) gives rise to a value \( d_i' \), then the overall value of \( d' \) is given by

\[
d' = \sqrt{\sum d_i'^2}.
\]

3. The value of \( d_i' \) is proportional to the square of the effective modulation index, \( m_i \), in the \( i \)th critical band:

\[
d_i' = km_i^2,
\]

where \( k \) is a constant of proportionality. This assumption is supported by the results of experiments 2 and 3.

For small depths of modulation, the difference in level between a maximum and minimum in excitation level, \( \Delta L \), is directly proportional to \( m_i \) \[ \Delta L_i = 20 \log_{10} \left( \frac{1 + m_i}{1 - m_i} \right) \approx 40 \log_{10} e m_i = 17.37m_i \]. Hence, for small \( m \),

\[
d_i' = c\Delta L_i^2
\]

where \( c \) is a constant. From Eqs. (8) and (10),

\[
d' = K \sqrt{\sum \Delta L_i^2},
\]

where \( K \) is a constant.

Excitation patterns and changes in excitation level were calculated using the program presented in Glasberg and Moore (1990). In this program, the excitation pattern is defined as the output from each auditory filter as a function of center frequency. Each auditory filter is assumed to have the form of a rounded exponential (Patterson et al., 1982):

\[
W(g) = (1 + pg) \exp(-pg),
\]

where \( W(g) \) is the intensity-weighting function defining the filter shape, \( g \) is the deviation from the center frequency of the filter divided by the center frequency, and \( p \) is a parameter determining the sharpness of the filter. The value of \( p \) can be different for the lower and upper halves of the filter; these values will be designated by \( p_l \) and \( p_u \), respectively. The equivalent rectangular bandwidth (ERB) of the filter, expressed as a proportion of center frequency, is \( 2/\rho_l + 2/\rho_u \). In the version of the program presented by Glasberg and Moore (1990), the value of \( \rho_l \) decreases with increasing level, while the value of \( \rho_u \) is invariant with level.

Initially, the model was applied to the psychometric functions for detecting AM alone and FM alone, as determined in experiment 3. The modulation indices required for \( d' = 1 \) were calculated from the fitted functions shown in Fig. 8. These indices were used to calculate the values of \( \Delta \), and \( \Delta \), for the AM stimuli, \( L_u \) and \( L_l \), and for the FM stimuli, \( F_u \) and \( F_l \), for all three subjects, \( K \) should have been equal for AM and FM. In fact the values were substantially different; for all three subjects, \( K \) was markedly greater for FM than for AM.

The discrepancy arose in the following way. The changes in excitation level produced by the FM were always small \(( <0.89 \text{ dB}) \) and were larger than 0.5 dB only over a very restricted region on the low-frequency side of the excitation pattern. Changes on the high-frequency side were always smaller than 0.34 dB. In contrast, the changes produced by AM were larger (up to 1.84 dB on the high-frequency side, and never less than 0.86 dB), and occurred over the whole of the excitation pattern. The large changes on the high-frequency side are a consequence of the nonlinear growth of excitation (changes in \( p_i \) with level) that are a feature of the program used (Glasberg and Moore, 1990).

One problem with the excitation-pattern model is that the filter shapes used are derived from simultaneous masking data. There is evidence that non-simultaneous masking can give a more accurate indication of the internal representation of stimuli (Houtgast, 1974; Moore and O’Loughlin, 1986). Auditory filters measured in forward masking have steeper high-frequency sides (greater values of \( p_u \) than
measured in simultaneous masking (Moore et al., 1987), and they do not vary substantially with level (Glasberg and Moore, 1982). Thus we decided to evaluate a version of the excitation-pattern model where the values of $p_t$ and $p_u$ did not vary with level, and with a greater value of $p_t$ than assumed in the program of Glasberg and Moore (1990). It was found that if $p_t$ was fixed at 25 and $p_u$ was fixed at between 30 and 56 (depending on the subject), the values of $K$ calculated from Eq. (11) were essentially the same for both AM and FM; the values of $K$ varied from 0.21 to 0.45, depending on the subject. These values of $p_t$ and $p_u$ are comparable to those found by Moore et al. (1987) for forward masking; their mean values for three subjects were 25 and 41 for $p_t$ and $p_u$, respectively. Thus it is possible to modify the excitation-pattern model in a plausible way so as to make it consistent with the detectability of both AM alone and FM alone.

Next, we determined whether the modified multiband excitation-pattern model could account for the values of $d'$ obtained for combinations of AM and FM. Assume that, for a specific value of $d'$, the level for AM alone varied between $L_i$ and $L_u$ and the frequency for FM alone varied between $F_i$ and $F_u$. Cases corresponding to each of the four modulator phases are described below.

Phase 0: A maximum in frequency coincides with a maximum in level. Hence, the values of $\Delta L_i$ were calculated as the differences in excitation level for a sinusoid with frequency $F_i$ and level $L_i$, and a second sinusoid with frequency $F_u$ and level $L_u$.

Phase $\pi$: A maximum in frequency coincides with a minimum in level. Hence, the values of $\Delta L_i$ were calculated as the differences in excitation level for a sinusoid with frequency $F_u$ and level $L_i$, and a second sinusoid with frequency $F_i$ and level $L_u$.

Phase $\pi/2$: The maxima and minima in frequency occur when the amplitude is at its mean value. The smallest and largest excitation levels on the low-frequency side of the excitation pattern occur approximately at these points. Hence, the values of $\Delta L_i$ for the low-frequency side of the pattern were calculated as the differences in excitation level for a sinusoid with frequency $F_i$ and a sinusoid with frequency $F_u$, both with a level of 70 dB SPL. The maxima and minima in level occur when the frequency is at its mean value. The smallest and largest excitation levels on the high-frequency side of the excitation pattern occur approximately at these points. Hence, the values of $\Delta L_i$ for the high-frequency side of the pattern were calculated as the differences in excitation level for a sinusoid with level $L_i$ and a sinusoid with level $L_u$, both with frequency 1000 Hz.

Phase $3\pi/2$: The calculations here are essentially the same as for the $\pi/2$ condition.

The predictions were similar for all three subjects. The model predicted that $d'$ values for combined AM and FM should be greatest for $\Delta \phi = \pi$, intermediate for $\Delta \phi = 0$, and smallest for $\Delta \phi = \pi/2$ and $3\pi/2$. Typically, $d'$ was predicted to be more than twice as large for $\Delta \phi = \pi$ as for $\Delta \phi = \pi/2$ and $3\pi/2$. The data do not conform to these predictions. The $d'$ values for $\Delta \phi = \pi$ tended to be lower than those for the other conditions rather than higher. Thus the multiband excitation-pattern model fails to account for the data.

### 3. Overview

The analysis presented above indicates that single-mechanism models based on excitation patterns do not account for the data; they predict phase effects that are inconsistent with the data. On the other hand, the results are also not consistent with the hypothesis that AM and FM are coded by completely independent mechanisms; the detection of MM at low rates was better than predicted from this hypothesis. We do not have a definite explanation for these results, but can suggest some possibilities.

FM may be coded both in terms of changes in excitation level and by some other mechanism, for example by changes in the pattern of phase locking evoked by the stimulus. AM may be detected primarily using changes in excitation level. When AM and FM coexist, changes in excitation level produced by the AM will combine with changes produced by the FM, giving performance somewhat better than for AM alone. However, the independent information provided by the second mechanism for FM will lead to a further improvement in performance.

Alternatively, it is possible that AM and FM are coded separately in the auditory periphery, but information about AM and FM is combined at some higher level in the auditory system, perhaps in channels tuned for detecting modulation. The ability to detect modulation might be determined by limitations in these channels, rather than by limitations in peripheral coding. If modulator phase is preserved at the inputs to these channels, then this could account for the phase effects observed in experiment 3, and for the finding that $d'$ for combined AM and FM was greater than predicted from the combination of independent sources of information.

### B. Detection of MM at high modulation rates: The Hartmann–Hnath model

According to the Hartmann–Hnath model, the detection of modulation at high modulation frequencies depends on the spectral structure of the modulated signal; performance is determined by the detectability of the lower sideband of the spectrum. The upper sideband, in Hartmann and Hnath’s opinion, is completely masked by the spectral component at the carrier frequency, and does not influence the modulation detection threshold. Assuming that threshold corresponds to a certain amplitude value of the lower sideband, $A_t$, the thresholds for detecting FM in the presence of a fixed amount of AM can be predicted using the formulas in Fig. 1(a)–(d). These formulas can be rearranged to give

for $\Delta \phi = 0$, $\beta = 2A_t/A_0 - m$, \hspace{1cm} (13)

for $\Delta \phi = \pi$, $\beta = 2A_t/A_0 + m$, \hspace{1cm} (14)

for $\Delta \phi = \pi/2$ and $3\pi/2$, $\beta = (2A_t/A_0)^2 - m^2$. \hspace{1cm} (15)

Figure 10 shows the mean data for the three subjects with $\omega_m = 256$ Hz. Since thresholds in this figure are expressed relative to the threshold obtained for FM alone ($\beta_{th}$), the constant terms in Eqs. (13)–(15) can be assigned the value 1. The dotted and dashed curves in Fig. 10 show the thresholds predicted using Eqs. (13)–(15). The pattern of
results follows the predicted values quite well (the points at 100% on the abscissa should be ignored, for reasons discussed earlier). The apparent small systematic errors for \( \Delta \phi = 0 \) and \( \pi \) could be due to errors in the measurement of the threshold for FM alone. Overall, then, the results are consistent with the Hartmann–Hnath model.

VII. CONCLUSIONS

The following conclusions can be drawn from our results.

1. Thresholds for detecting FM depended significantly on coexisting "subthreshold" AM.

2. At modulation frequencies below 64 Hz, thresholds (corresponding to \( d' \approx 1.16 \)) did not depend significantly on the relative phase of the FM and AM modulators (experiment 1). However, the detectability of combined AM and FM using a 10-Hz modulation frequency did depend on relative modulator phase for higher values of \( d' \) (experiment 3). Detectability was highest for relative modulator phases of 0 and \( \pi/2 \), and lowest for a relative modulator phase of \( \pi \).

3. At low modulation rates (<16 Hz), the detectability of AM alone or FM alone, as measured by the detectability index \( d' \), was proportional to the square of the modulation index, for both AM and FM.

4. The results for the detectability of combined AM and FM could not be explained by a single-band excitation-pattern model or a multiband excitation-pattern model.

5. The results are also inconsistent with a model based on the assumption that AM and FM are detected using completely independent mechanisms; the detectability of combined AM and FM was higher than predicted from this assumption.

6. At a high modulation rate (256 Hz), thresholds for detecting FM in the presence of "subthreshold" AM showed a strong dependence on relative modulator phase. These results could be explained by the Hartmann–Hnath model which assumes that detection of modulation depends exclusively on the amplitude of the lower sideband in the spectrum of the modulated signal.

ACKNOWLEDGMENTS

We would like to express our gratitude to the Cambridge Hospitality Scheme Committee for providing funds to finance A. Sek's stay in Cambridge. We would like also to thank Brian Glasberg for writing the computer programs used in these experiments. Thomas Baer, Brian Glasberg, Michael Shailer, and Michael Stone provided helpful comments on an earlier version of this paper. Brian Buck and Helen Morley helped to gather the data for experiment 3. This work was supported, in part, by the State Committee for Scientific Research P. No. 209109101, and by the MRC.


