PURE-TONE OCTAVE MASKING IN NORMAL-HEARING LISTENERS

DAVID A. NELSON *and* ROBERT C. BILGER

University of Pittsburgh, Pittsburgh, Pennsylvania

Octave masking was investigated at four different frequencies (250, 500, 1000, and 2000 Hz) as a function of intensity of the masker and phase of the test signal. Slopes of phase-locked octave masking were found to increase with masking signal frequency, from 0.80 dB/dB at 250 Hz to 3.0 dB/dB at 2000 Hz. The monaural octave-masking phase effect was considerably larger for masking signals at low frequencies than at high frequencies, and the phase effect decreased or disappeared entirely for high-level masking signals. Interpretations are considered which take recent neurophysiological and physiological data into account, and which describe the octave-masking phase effects in terms of temporal pattern discrimination. Those interpretations adequately account for the frequency dependencies found in octave-masking phase effects.

The perception of subjective harmonics of single pure tones and the perception of interactions among subjective harmonics of multiple tones led early investigators of audition to hypothesize that the auditory system is nonlinear and therefore generates distortion components that are not contained in the acoustic signal (Boring, 1942, pp. 357-359). The asymmetry of masked audiograms for pure-tone maskers (Wegel and Lane, 1924; Fletcher, 1929, pp. 167-187; Egan and Hake, 1950; Small, 1959) and for narrow-band noise maskers (Egan and Hake, 1950; Bilger and Hirsh, 1956) has been interpreted as further evidence that distortion components exist in normal ears and that those distortion components are capable of masking test signals at the frequencies of the distortion components. More recently, Kameoka and Kuriyagawa (1966) and Clack (1967, 1968) have studied phase-locked octave masking (tone-on-tone masking) using the steady-tone technique. Monaural phase effects have also been investigated by Beasley (1930), Craig and Jeffress (1962), Fricke (1968), and Raiford and Schubert (1971). Consideration of the results from those octave-tone experiments, however, presents an incomplete picture.

One of the goals of contemporary psychoacoustics is to understand pathological hearing. Results to date, however, do not provide the data on normal ears that are necessary for valid comparisons with octave-masking data from sensorineural ears. The hearing losses exhibited by sensorineural ears necessitate the use of a wide range of signal intensities and a wide range of signal frequencies. Clack (1968) has obtained octave-masking data from normal

ears, but not for high intensity levels, nor for low as well as high frequencies. To arrive at a set of octave-masking data from normal ears, a series of experiments was carried out to definitively describe intensity, frequency, and phase effects in two "laboratory" ears. From those data, variables were chosen that were likely to provide reliable information about phase-locked octave masking from inexperienced listeners in a realistic amount of test time. Then a group of normal-hearing listeners with no listening experience was tested to yield a set of octave-masking data to be used for comparison with similarly obtained octave-masking data from a population of outpatient listeners with sensorineural hearing loss. The investigations of normal-hearing listeners are reported here, along with discussion of relevant explanatory models. The investigation on sensorineural ears is reported in a companion paper (Nelson and Bilger, 1974).

METHOD

Apparatus

A block diagram of the experimental apparatus used in these experiments is shown in Figure 1. Masking signals (S_m) at four different frequencies (f_1)

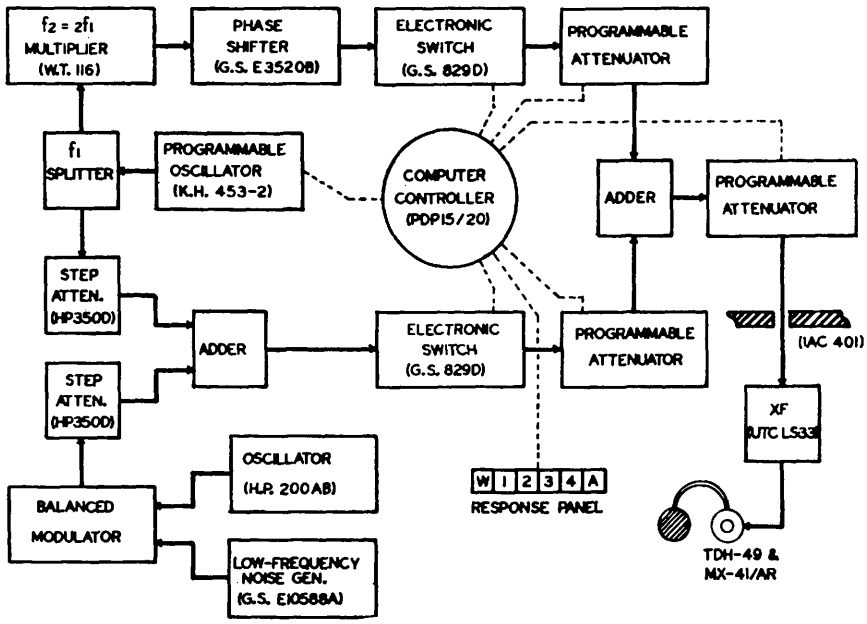

FIGURE 1. Block diagram of the equipment used in phase-locked octave masking.

were used (250, 500, 1000, and 2000 Hz). A function generator (Wavetek Model 116) doubled f_1 to yield test signals (S_t) with frequencies of $f_2 = 2f_1$. Control over the phase between S_m and S_t was maintained by passing S_t

through a phase shifter (Grason-Stadler Model E3520B). During some of the experiments, it was necessary to mix a narrow-band noise (NBN) with the masking signal in order to mask aural-distortion components that were not part of the experiment. The narrow-band noise was produced by balance modulating (double-side band, suppressed carrier) a carrier frequency (f_c) with a noise band from 5 to 200 Hz. That low-frequency noise band had a high-frequency slope of 40 dB per octave. The bandwidth of the resulting noise was 400 Hz. Both the high-frequency and the low-frequency slopes of this band were 40 dB per 200 Hz. Its center frequency was equal to the carrier frequency (NBN = $f_c \pm 200$ Hz). The masking signals at f_1 (plus narrow-band noise when applicable) and the test signals at f_2 were gated simultaneously with a pair of electronic switches that were controlled by the computer. The level of each tone was separately controlled by two programmable attenuators (Wolf, 1972), then S_m and S_t were added together. The level of the combined wave form was controlled by a third programmable attenuator.

To specify when a rarefaction or condensation wave was delivered to the ear canal, the polarity of the acoustic output of the earphone was determined by observation of an earphone diaphragm under a high-power operating microscope.

The phase shift between the electrical and acoustic wave forms (measured with a condensor microphone in a 6-cc coupler) was less than 7° for S_t relative to S_m. Phase relations in these experiments are expressed either in degrees of phase shift of S_t relative to S_m (or in radians). The general trigonometric expression for the waveform is:

$$A \sin \omega t + B \sin (2\omega t + \phi)$$

To avoid ambiguity, both conventions for specifying phase relations between S_m and S_t are shown in Figure 2 along with the appropriate wave forms for equal amplitude tones at frequencies of f_1 and f_2. Positive values in Figure 2 indicate inward displacement of the eardrum. Phase calibration throughout these experiments was performed oscilloscopically with a 90° ($\phi = \pi/2$) $S_m + S_t$ acoustic waveform.

Subjects

The initial experiments to define intensity, frequency, and phase effects in octave masking (Experiments I, II, and III) used a single ear from each of two normal-hearing (ANSI, 1969) experienced listeners (SBH and DAN). Since one of the purposes of these experiments was to develop techniques for obtaining octave-masking data from relatively inexperienced outpatient listeners with hearing loss, normal-hearing listeners, who had never participated in this type of listening task before, were chosen for the group experiment (Experiment IV). A total of 12 normal-hearing listeners was used in the group experiment. Of the 12, six were outpatients from the audiology clinic

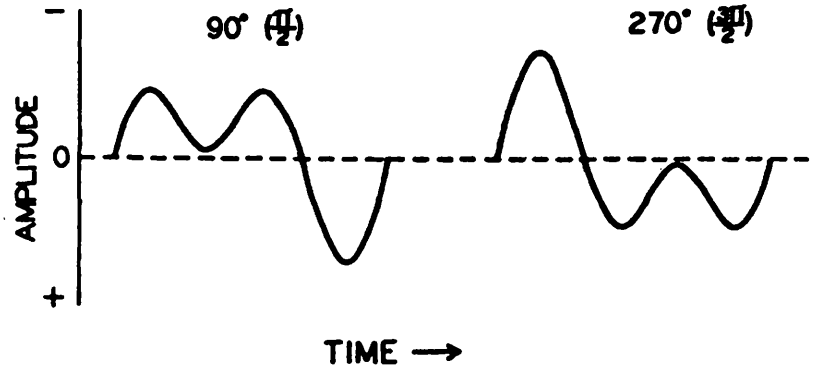

FIGURE 2. Amplitude waveforms for the 90° ($\pi/2$) and the 270° ($3\pi/2$) $f_1 + 2f_1$ signals. Both conventions (degrees of phase shift and radians) for specifying phase relations of the $2f_1$ test signal relative to the f_1 masking signal are shown. Phase is specified in terms of the period of the $2f_1$ test signal. The trigonometric expression for these waveforms is $A \sin \omega t + B \sin (\omega t + \phi)$, where $\phi = \pi/2$ and $3\pi/2$.

who had normal hearing in the test ear at the frequencies employed in the experiments. Four of the 12 were nonlaboratory personnel who had normal hearing. The remaining two normal-hearing listeners were DAN and SBH, the trained listeners.

PROCEDURES

Throughout the experiments, a four-interval forced-choice (4IFC) adaptive psychophysical procedure was employed to determine the intensity ratio between S_m and S_t at which 50% correct discriminations between S_m alone and $S_m + S_t$ could be made (the threshold for S_t when masked by S_m). Each listener was seated in a sound-treated room in front of a panel containing a warning light, four observe lights, and an answer light. As illustrated in Figure 3, a listener's task was to listen to the sounds that coincided with each of the four observe lights and, when the answer light came on, to push the button corresponding to the interval in which they had heard a sound that was different (in any way) from the sounds in the other three intervals. The f_1 masker occurred in all four intervals, and the f_2 test signal occurred in only one interval, determined randomly from trial to trial. The time periods shown in Figure 3 were maintained during the group experiment and during the experiments on listener SBH. Because of the suggestions of Nixon, Raiford, and Schubert (1970), the interval between signals was shortened from 250 to 100 msec for some of the experiments performed on DAN. As in their experiment, shortening the off interval improved DAN's performance and made the discrimination tasks subjectively easier.

In the "adaptive" psychophysical procedure, each threshold determination (run) was started at an S_m/S_t intensity ratio that was easy to discriminate.

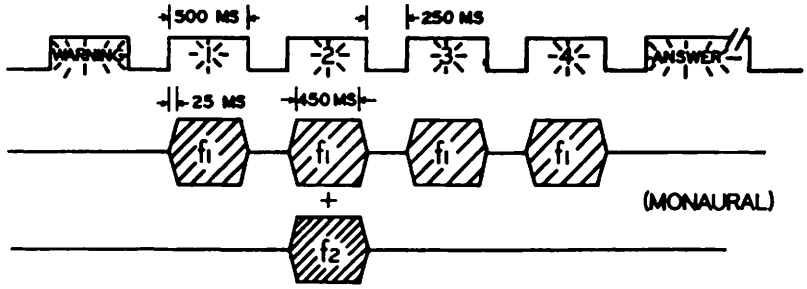

FIGURE 3. Graphic representation of the temporal parameters that existed in the four-interval forced-choice paradigm.

When a listener correctly discriminated two consecutive 4IFC trials at a given S_m/S_t intensity ratio, the level of S_t was lowered by 2 dB (accept). When a listener missed both 4IFC trials, the level of S_t was raised by 2 dB (reject), and when a listener missed one of a pair of 4IFC trials (defer), two more 4IFC trials were presented at that same S_t level. After three reversals in the direction of level changes, or after five consecutive defers, the run automatically terminated. That level at which a listener responded correctly on 50% of the trials was taken as threshold. In some instances, it was necessary to interpolate to the 1-dB midpoint between two consecutive 2-dB steps. A single run took two to four minutes. An experimental session lasted 40 to 50 minutes.

During a single experimental session, only one pair of frequencies was tested (f_1 and f_2). First, a quiet threshold was obtained for $S_m(f_1)$ alone and for $S_t(f_2)$ alone. Then eight to 12 thresholds for S_t masked by S_m were determined. Finally, quiet thresholds for S_m and S_t were repeated. The earphone was not removed until the end of a session. When level effects were being investigated, maskers with an S_m-S_t phase of either 90° or 270° were usually presented in 10-dB steps from 10 to 100 dB SPL, beginning with 10 dB SPL. When phase effects were investigated, the level of S_m was held constant and the phase angle between S_m and S_t was varied from 0° through 360°, usually in 45° steps. In experiments on intensity effects and phase effects performed with SBH and DAN, no significant order effects were found, so the remainder of the data was collected with the constant orders of conditions stated above.

Clack (1967, 1968) reported considerable variability among retests in phase-locked octave-masking data. He attributed some of the variability to earphone placement and movement of the jaw. During informal listening sessions by one of us (DAN), differences in repeated S_t masked threshold at f_2 were found to be larger than 10 dB. Such differences could be produced consistently by tilting the head from one side to the other, especially when low-frequency signals were employed as maskers. Those large changes in S_t masked thresholds probably were due to small leaks in the ear/earphone seal, resulting in small level changes in S_m that produced large changes in S_t masked thresholds. That result is a reasonable expectation, since some octave-

masking functions have steep slopes, sometimes as steep as 3 dB/dB. Consequently, all listeners in these experiments were instructed to maintain a constant head position while listening and were instructed not to touch the earphones once the earphones had been placed on their ears by the experimenter.

RESULTS

Octave-Masking Functions (Experiment I)

Preliminary listening experiments by DAN and SBH demonstrated that maximum masking occurred with a phase angle of 90° ($\pi/2$) between S_m and S_t, and minimum masking occurred with a phase angle of 270° ($3\pi/2$). Since Clack's (1968) data showed no shift in the phases corresponding to maximum and minimum masking when S_m level was varied, we decided to obtain octave-masking functions from SBH and DAN using the 90° and 270° phase conditions. The purpose of Experiment I was to define level effects and frequency effects in phase-locked octave masking. Figure 4 summarizes the octave-masking data obtained from DAN for masking tones with frequencies of 250, 500, 1000, and 2000 Hz at phase angles of 90° and 270°. Comparable data from SBH is shown in Figure 5.

Two sets of coordinates are included in Figures 4 and 5. The level of the masking signal (S_m) is given on the abscissa in sensation level (f_1 dB SL) at the bottom of each graph, and as sound pressure level (f_1 dB SPL) at the top of each graph. Similarly, masking of the test signal (S_t) by S_m is given on the ordinate as amount of masking (f_2 dB SL) at the left of each graph, and as masked thresholds (f_2 dB SPL) at the right of each graph. Data obtained with the 90°-phase conditions are represented by filled circles, and data obtained with the 270°-phase conditions are represented by unfilled squares. The data points are the means of three to six threshold determinations. A few data points are based on only one or two threshold runs, since only a check was desired at those conditions. The vertical lines through the symbols represent one standard deviation above and below each mean. The remaining symbols in Figures 4 and 5 are part of other experiments and will be described later.

Examination of Figures 4 and 5 shows that phase-locked octave masking produced by a pure tone was basically nonlinear, as has been shown in the past by Wegel and Lane (1924) for tones higher in frequency than the pure-tone masker, and by Chapin and Firestone (1934) for phase-locked tones an octave apart. At low levels of S_m, little or no masking of S_t occurs. At moderate levels of S_m, masking of S_t grew rapidly. At high levels of S_m, the masking functions flattened out, that is, above about 80 dB SPL a 10-dB change in the level of S_m resulted in a change of 10 dB or less in S_t masked threshold. The 90°-phase conditions yielded maximum masking and the 270°-phase conditions yielded minimum masking.

Comparison of the masking functions of DAN in Figure 4 with the masking

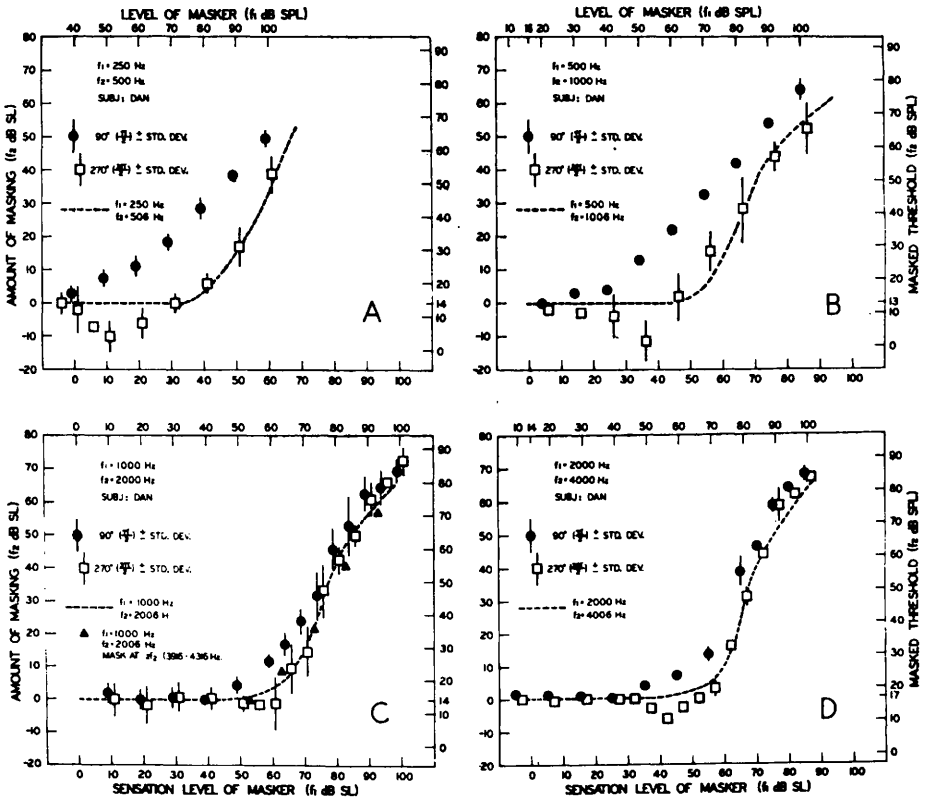

FIGURE 4. Phase-locked octave-masking functions from listener DAN for four different f_1 masking signals (Experiment I). Data points are mean masked thresholds for $f_2 = 2f_1$ test signals masked by f_1 masking signals. Masking functions for the 90° (cancellation) phase, which produced maximum masking, are shown as filled circles. Masking functions for the 270° (augmentation) phase, which produced minimum masking, are shown as open squares. Waveforms corresponding to the 90° and 270° phases are shown in Figure 2. Mistuned octave-masking functions for $f_2 + 6$ Hz test signals are shown by the dashed lines (Experiment III). The filled triangles in 2C show a mistuned masking function obtained while the second harmonic (at $2f_2$) of the f_2 test signal was masked. Coordinates for all four graphs are given in sound pressure level (SPL) and in sensation level (SL).

functions of SBH in Figure 5 shows several differences between the two listeners. The differences between the 90°- and 270°-masking functions were larger for DAN than for SBH. Listener SBH also showed more saturation of the masking functions at high levels of S_m than did DAN. These differences are mentioned here to call attention to the fact that individual differences in octave masking are large (Clack, 1968), that the listening tasks discussed here are complex, and that the sources of these individual differences are not clearly understood. Perhaps these listener differences were real, that is, they were physiological, or perhaps they were solely a function of listening experience. Since listener DAN's results were much more reliable than SBH's results, most of the remaining experiments were limited to DAN alone.

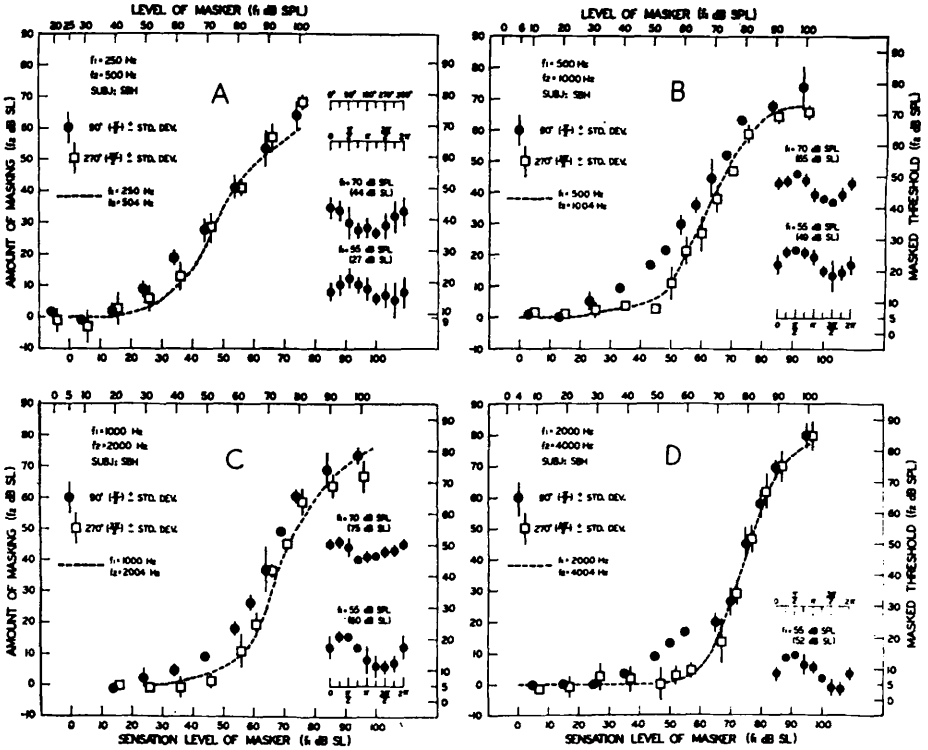

FIGURE 5. Octave-masking functions obtained from the less experienced listener, SBH (Experiment I). Data points are mean masked thresholds for $f_2 = 2f_1$ test signals masked by f_1 masking signals. Masking functions for the 90° (cancellation) phase, which produced maximum masking, are shown as filled circles. Masking functions for the 270° (augmentation) phase, which produced minimum masking, are shown as open squares. Waveforms corresponding to the 90° and 270° phases are shown in Figure 2. Mistuned octave-masking functions for f_2 + 6 Hz test signals are shown by the dashed lines (Experiment III). The filled triangles in 2C show a mistuned masking function obtained while the second harmonic (at $2f_2$) of the f_2 test signal was masked. Coordinates for all four graphs are given in sound pressure level (SPL) and in sensation level (SL). Masked thresholds for f_2 test signals are shown as a function of phase in the right-hand insets (Experiment II). The phase relations between the f_1 masking signal and the f_2 test signal ($f_2 + 2f_1$) are specified on the abscissa in terms of one period of the f_2 test signal. Phase is given in radians, and for convenience, it is also given in degrees in 5A.

An unexpected result, not seen in previous octave-masking data, was the occurrence of large differences among octave-masking functions for different S_m frequencies. Those frequency effects were apparent in the masking functions of both listeners and are perhaps best expressed as differences in the slopes of the octave-masking functions. Figure 6 shows the relation between S_m frequency (f_1) and the slopes of the steepest portion of the masking functions of Figures 4 and 5. The data are fit by eye with a dashed line to show the trend towards steeper masking slopes as S_m frequency increases from 250 to 2000 Hz. Masking slopes varied from less than 1 at 250 Hz to greater than

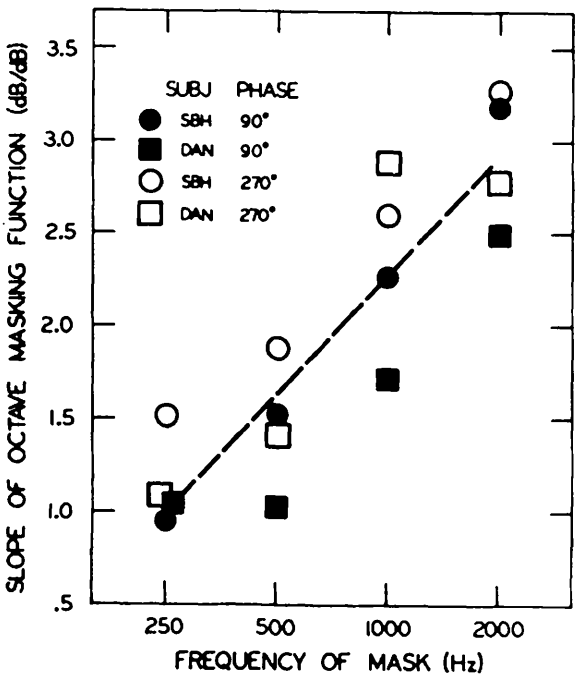

FIGURE 6. Slopes of octave masking as a function of the frequency of the f_1 masking signals. Slopes, in dB of masking per dB of masker increase, were taken from the steep portions of the phase-locked octave-masking functions shown in Figures 4 and 5. The dashed line, a visual fit, has a slope of about 0.65 dB/dB for each octave increase in the masking tone.

3 at 2000 Hz. The average increase in masking slope was about 0.65 dB/dB for each octave increase in masking frequency. Listener DAN tended to show slopes that were less steep than slopes shown by SBH. Slopes for the 270°-phase conditions were steeper than for the 90°-phase conditions.

The differences between 270°- and 90°-masking phases were also larger at low S_m frequencies than at high. This result can be seen in DAN's data in Figure 4. Since the two phase conditions were the phases that produced maximum masking and minimum masking, these results suggest that there are larger monaural phase effects at low S_m frequencies than at high S_m frequencies. However, Clack (1968) reported a peculiar phase reversal at moderate levels of S_m in one of his listeners, and since only two phases were measured in the functions shown in Figures 4 and 5, further investigation of phase effects as a function of intensity and frequency was necessary.

Monaural Phase Effects (Experiment II)

Masked thresholds for S_t were obtained from DAN and SBH as a function of the phase between S_m and S_t at various levels of S_m and at each of the four S_m frequencies. The purpose was to look for possible phase reversals at high levels of S_m, and to determine the size of the phase effect as a function of S_m frequency and S_m intensity. The results of Experiment II for listener DAN are shown in Figure 7. The results of Experiment II for listener SBH are shown in the right-hand insets of Figure 5. The amount of masking of S_t is shown on

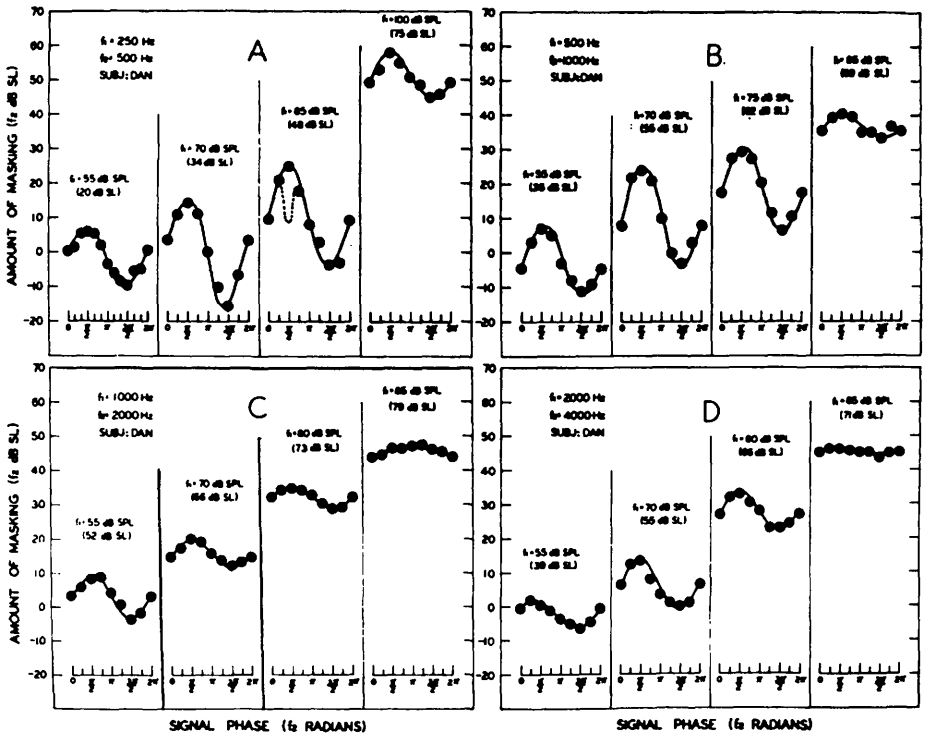

FIGURE 7. Octave masking as a function of phase for four different f_1 masking signals at four different masker levels from listener DAN (Experiment II). The phase relations between the f_1 masking signal and the f_2 test signal ($f_2 = 2f_1$) are specified on the abscissa in terms of one period of f_2. See text for explanation of dashed line in 7A.

the ordinates (f_2 dB SL). The phase angles between S_m and S_t are shown on the abscissa in radians. Scales specifying phase in both degrees and radians are shown in Figure 5A for reference. These data show the classical sinusoidal change in masking that occurs when the phase relations between S_m and S_t are varied (Chapin and Firestone, 1934).

A phase reversal as a function of S_m level would be seen as a change in the phases at which maximum and minimum masking occur when S_m level is varied. Only for listener SBH did such a phase reversal appear (Figure 5A), and then only at 70 db SPL with a masker of 250 Hz, and with considerable variability relative to the small phase effects SBH exhibited. In all the phase functions for DAN (Figure 7), maximum masking occurred with a phase angle of 90° and minimum masking occurred with a phase angle of 270°. When S_m was increased from low SLs, the phase effect increased to a maximum, and then became smaller, sometimes disappearing at high levels of S_m. There were no phase reversals for listener DAN.

The magnitude of these monaural phase effects can be seen for listener DAN (Figure 7) largely as a function of S_m frequency (f_1). The greatest difference between maximum and minimum masking was as large as 30 dB for low-

frequency maskers, and no more than 12 to 14 dB for high-frequency maskers. Comparison of the magnitude of the phase effect among S_m frequencies at comparable S_m levels in Figure 7 (either in terms of dB SPL or dB SL), clearly shows larger phase effects at 250 and 500 Hz than at 1000 and 2000 Hz.

Some of the difficulties encountered in performing pure-tone masking experiments were demonstrated in the results of DAN for a masker at 250 Hz. The phase effect for S_m at 85 dB SPL in Figure 7A showed an unusual notch at 90°. That notch is illustrated by the dashed line. When only 45°-phase increments were tested, the 90° condition showed considerable test-retest variability. However, when consecutive 15°-phase increments were tested, and when considerable care was taken to maintain a constant ear/earphone position, the notch shown in Figure 7A was reliably observed. We hypothesized that off-frequency listening (that is, detection at some frequency region other than f_2) might account for this unexplained notch. Repetition of that phase function in 15°-phase increments and in the presence of a narrow band of masking noise in the frequency region of $2f_2$ (900-1300 Hz, slopes better than 80 dB/octave, at 55-dB spectrum level) completely obliterated the 90° notch but did not mask any other portion of the phase function. The 90° notch, therefore, was explained as the detection of interference from the second harmonic (at $2f_2$) of the test signal instead of detection of the test signal (at f_2) itself.

Mistuned vs Phase-Locked Octave Masking (Experiment III)

Since the goal of all of these investigations was to obtain phase-locked octave masking from normal ears for comparison with similar data from sensorineural ears, it was of value for us to determine how a technique previously used for estimating aural distortion (Wegel and Lane, 1924; Newman, Stevens, and Davis, 1937; Clack and Bess, 1969), referred to here as mistuned-octave masking, compared with phase-locked octave masking. Clack and Bess (1969) have obtained mistuned-octave-masking data from listeners with sensorineural hearing loss. We, therefore, wanted normal data with which to compare mistuned- and phase-locked octave masking. The mistuned-octave-masking technique utilizes a test signal (S_t) that has a frequency about 4 Hz higher than twice the frequency of S_m. S_t is allowed to beat with the second harmonic of S_m at a rate of four times per second. A masked threshold is then obtained (when $f_2 = 2f_1 + 4Hz$) in the presence of different levels of S_m at a frequency of f_1.

Test signals with a frequency of f_2 were mistuned by 4 Hz for listener SBH and by 6 Hz for listener DAN ($f_2 = 2f_1 + 6Hz$). By listening to the beats between a signal at f_2 and a signal at $f_2 + 6Hz$, listener DAN determined that the beats he heard during the mistuned-masking experiments were actually at a rate of 6 Hz. Masking functions for mistuned octaves were obtained at all four S_m frequencies from both listeners. The results are shown as dashed curves in Figures 4 and 5. At low levels of S_m, the mistuned masking functions

lie between the 90°- and the 270°-phase-locked functions, but only when there is a large phase effect that pushes the 270° function below the S_t quiet threshold. This occurred for DAN at all S_m frequencies. Listener SBH's 270° functions never did drop substantially below S_t quiet threshold. Her mistuned functions fall almost on top of her 270° functions at low S_m levels. At moderate and at high S_m levels, the mistuned functions for both listeners tend to fall on top of the 270° phase-locked functions. From these results, we can conclude that mistuned-octave-masking functions correspond closely to phase-locked octave-masking functions that are obtained with a minimum masking phase, except at low S_m sensation levels when the phase-locked function drops below the threshold for S_t alone.

Just as it was hypothesized earlier that the detection of a distortion product of the test signal in the frequency region of $2f_2$ might influence phase functions, we hypothesized that the flattening of the steep octave-masking function at high S_m levels might also be due to the detection of distortion products of S_t. Specifically, we suspected the second harmonic of S_t ($2f_2$). To test this possibility, a mistuned-octave-masking function was obtained from listener DAN at a masker frequency of $f_1 = 1000$ Hz, while a narrow-band noise (3916-4316 Hz, at a 55-dB spectrum level) was gated simultaneously with the masker to mask any distortion components at $2f_2$. The mistuned-octave-masking function that resulted is shown in Figure 4C by the filled triangles. Because there were no appreciable differences between that function and the regular mistuned masking function (shown by the dashed curve), detection of the second harmonic of S_t was not considered to be an important variable.

As an explanation for part of the nonlinearity of octave-masking functions, it occurred to us that instead of detecting an interaction between S_t and the "aural" harmonic of S_m, listeners may be detecting an interaction between S_t and the "acoustic" harmonic of S_m. Therefore, a portion of the octave-masking function may reflect the amount of test signal needed to just detect an amplitude increment (ΔI) at different levels of the masker and its acoustic second harmonic. To check this possibility, just discriminable difference limens were obtained from both listeners at various frequencies. This was done simply as a mistuned-masking experiment where the masker and test signal were about 6 Hz apart. Instead of plotting the data as difference limens, they were plotted on sensation-level coordinates as classical masking functions (Wegel and Lane, 1924). Figure 8 shows several mistuned masking functions from listener DAN for the middle frequencies. The frequency of the masker (S_m) is labeled f_1 and the frequency of the test signal (S_t) is labeled f_2. When S_m and S_t were close in frequency, the masking functions were linear with slopes less than one (between 0.77 and 0.93 for all the functions obtained from both listeners). The straight-line curves for $f_2 = 1006$ Hz and $f_2 = 2006$ Hz shown in Figure 8 were fit to the data with a least-squares procedure. In contrast to those two straight-line functions, mistuned masking of test signals at frequencies above the frequency of the pure-tone masker are also shown. (They will be discussed later.)

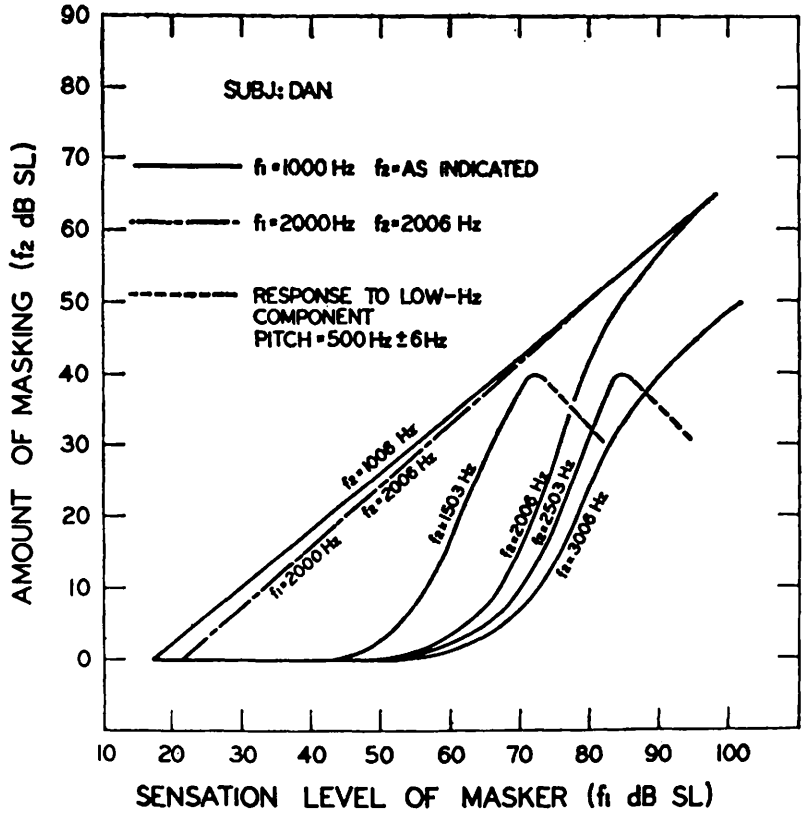

FIGURE 8. Mistuned pure-tone masking functions at f_2 = 1006, 1503, 2006, 2503, and 3006 Hz (Experiment III).

Given that the level of the acoustic second harmonic of S_m is more than 46 dB down from S_m, when f_1 = 1000 Hz, the expected contribution of the acoustic second harmonic of S_m can be calculated from the masking functions shown in Figure 8. For example, when S_m (at f_1 = 1000 Hz) was at 90 dB SPL, its acoustic second harmonic (at $2f_1$ = 2000 Hz) was less than 44 dB SPL. A 2000-Hz signal at 44 dB SPL was 30 dB above threshold, or 30 dB SL. From the data in Figure 8, we see that for a 2000-Hz masker at 30 dB SL, the masked threshold at 2006 Hz was 7 dB SL. Now, with a 1000-Hz masker at 90 dB SL, the masked threshold at 2006 Hz was 57 dB SL, 50 dB higher than would be predicted if the acoustic second harmonic of S_m (at $2f_1$ = 2000 Hz) were responsible for the masking. So, it appears that for high-level pure-tone maskers, the contribution of the acoustic second harmonic of S_m is negligible.

Similarly, at lower levels of S_m, say at 70 dB SL, one would predict an acoustic second harmonic at 2000 Hz of less than 26 dB SPL, or less than 12 dB SL. A 2000-Hz signal at 12 dB SL produced no masking at all at 2006 Hz. The masked threshold at 2006 Hz with a 70-dB-SL 2000-Hz masker was 15

dB SL, at least 15 dB higher than necessary if the acoustic second harmonic of S_m were contributing. So, at moderate levels of S_m, it is also unlikely that the acoustic second harmonic of S_m is a significant variable in octave masking.

The other functions in Figure 8 demonstrate comparable masking functions from listener DAN at the third harmonic of S_m ($f_2 = 3006$ Hz), at a frequency between S_m and the second harmonic of S_m ($f_2 = 1503$ Hz), and at a frequency between the second and third harmonic of S_m ($f_2 = 2503$ Hz). The dashed lines in Figure 8 reflect the results of detecting difference tones at a frequency of $f_2 - f_1$ for high levels of S_m. Figure 9 shows that masking with a narrow-band noise in the region of $f_2 - f_1$ difference tones eliminated the contamination by those difference tones, and shows that once $f_2 - f_1$ difference tones have been adequately masked, the masking functions between harmonic frequencies are similar to masking functions at harmonic frequencies.

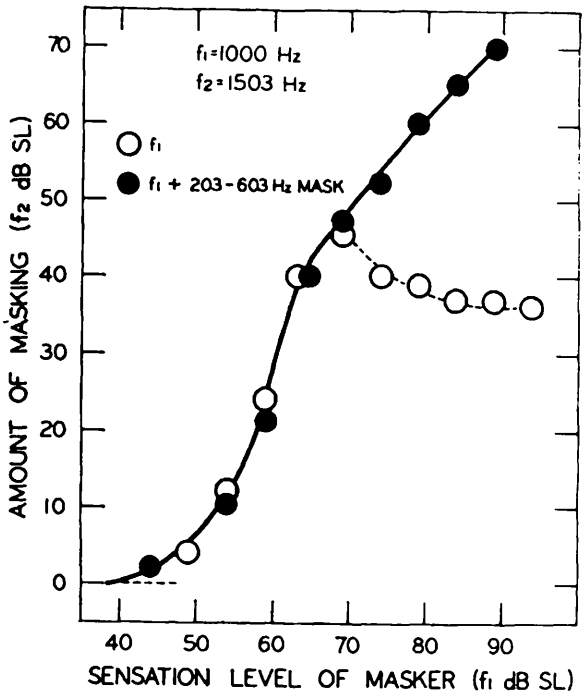

FIGURE 9. Mistuned masking functions for test signals at $f_2 = 1503$ Hz masked by masking signals at $f_1 = 1000$ Hz, with and without masking in the region of $f_2 - f_1 = 503$ Hz difference tones.

Normative Data from Inexperienced Listeners (Experiment IV)

The octave-masking data discussed above was used to select conditions for testing inexperienced normal-hearing listeners and listeners with sensorineural hearing loss. A phase setting of 90° was chosen because it yielded maximum masking in DAN and SBH as well as in some pilot-study outpatients whose data are not presented here. The 90°-phase condition also produced less variability from test to retest in DAN and SBH than did the 270° condition. This can be seen by comparing the vertical bars in Figure 4 and Figure 5 which indicate one standard deviation above and below each mean. For the

90° condition, a standard deviation was often as low as 2 dB while the comparable standard deviation for the 270° phase was sometimes over 10 dB. In the earlier experiments, both listener DAN and listener SBH reported it was more difficult to settle on a listening criterion during the 270°-phase conditions than during the 90° conditions.

The 90°-phase-locked octave-masking functions were obtained from 12 inexperienced listeners using the same psychophysical procedures employed with the experienced listeners. These were also the same psychophysical procedures to be used with outpatients who exhibit sensorineural hearing loss. These normative data are presented in Figure 10 for S_m frequencies at f_1 of 250, 500, 1000, and 2000 Hz. Masked thresholds for S_t are shown on the ordinate (f_2 dB SPL) as a function of the level of the masker (f_1 dB SPL). Sound pressure level coordinates are used here, since we felt that they provide the most valid comparisons among listeners with sensorineural hearing

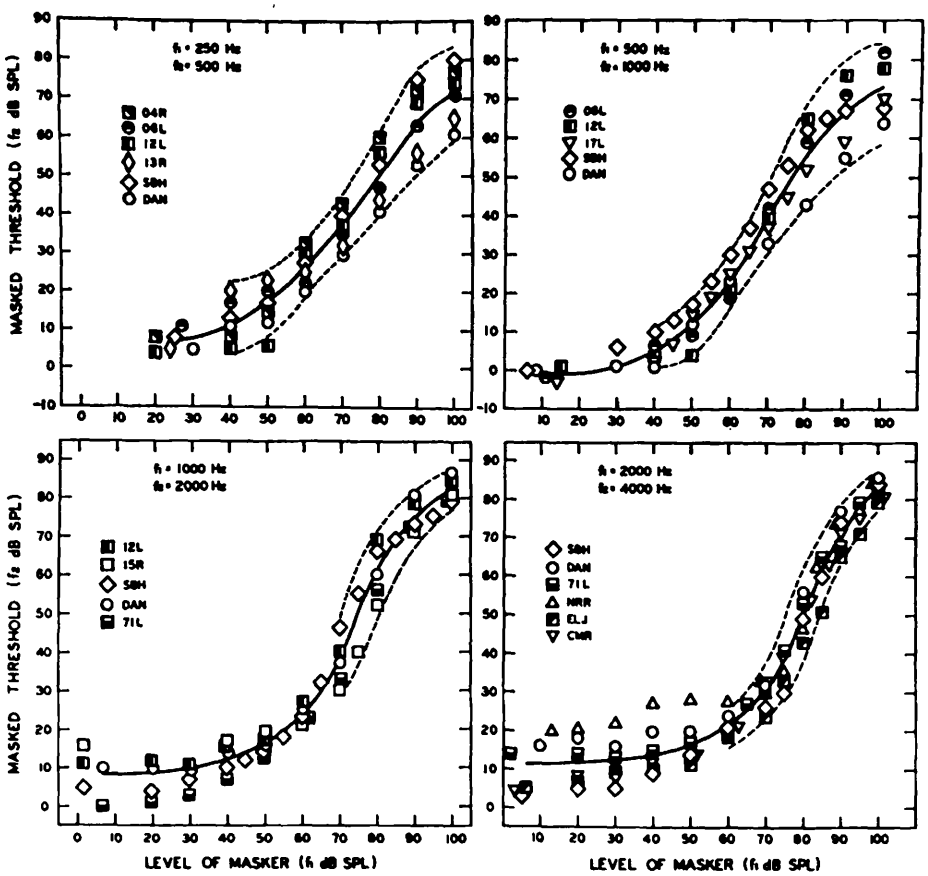

FIGURE 10. Octave-masking functions, phase-locked at 90° for four different f_1 masking signals, from normal-hearing inexperienced listeners (Experiment IV). The 90° waveform is shown in Figure 2.

loss (Nelson and Bilger, 1974). The solid curves are visual fits of the means at each S_m level. The dashed lines are the 85% confidence limits at each S_m level, and are intended to be used as estimates of the expected range for normal octave masking. The majority of the functions in Figure 10 are similar to those obtained earlier from SBH and DAN. The only noteworthy difference is that the inexperienced listeners exhibited octave masking which showed slightly more saturation or bend over at high levels of S_m than did DAN. There also tends to be less variability among listeners at high frequencies than at low frequencies. This result reflects small conductive losses for the more variable low frequencies in some of those normal ears. The sets of data in Figure 10 represent what one might expect from a large sample of normal ears and will be used in a companion paper (Nelson and Bilger, 1974) in a comparison with similar data from sensorineural ears.

DISCUSSION

Any attempts at explanatory modeling of aural distortion have to account for three basic generalizations from the experiments described above. First, the size of the monaural phase effect has been shown to be considerably larger for low- than for high-frequency maskers (S_m), a result we shall refer to as the frequency-dependent phase effect. Second, maximum and minimum masking phases were 90° and 270°, suggesting a 90°-phase difference from what would be predicted with traditional models of aural distortion. Third, the slope of octave masking has been shown to be a function of fundamental frequency, with steeper slopes for high frequencies, a result we shall refer to as frequency-dependent masking slopes. These three findings will be dealt with in the discussion that follows.

Harmonic Distortion and Vector Summation

One traditional interpretation of monaural phase effects and phase-locked octave-masking data relies heavily upon concepts of harmonic distortion and upon a vector-summation model. It is assumed that the auditory system, like most biological systems, is a nonlinear system, that is, that the output is not a constant proportion of the input. For example, quadratic nonlinearity is one specific type of nonlinearity in which the output is proportional to the square of the input ($y = ax^2$). Peak clipping is another type of nonlinearity in which the system exhibits saturation or amplitude limiting. The system is linear below a certain value of the input, but the output does not increase in proportion to the input once the value of that input is surpassed. The input-output characteristic for this saturation type of nonlinearity can be approximated by a polynomial equation of the form $y = a_0 + a_1x + a_2x^2 + \ldots a_nx^n$. In both types of nonlinearity, the output waveform is still periodic but is distorted relative to the input waveform. Customarily, distorted periodic output waveforms are described mathematically according to the theorem of Fourier, as

the sum of harmonically related sinusoids or as harmonic distortion components.

The concept of harmonic distortion implies energy in frequency regions not contained in the input signal, but a vector-summation model is needed to account for sinusoidal monaural phase effects. A vector-summation model simply describes the manner in which a presumed internal harmonic distortion component and an actual external probe tone interact. Supposedly, the auditory system detects the sum of an internally generated aural-harmonic distortion component (or aural harmonic) and an externally generated test signal of the same frequency. A vector-summation model predicts that the amplitudes of the two components of equal frequency will sum when the two components are "in phase" (0°) and that they will cancel when "out of phase" (180°). Since Craig and Jeffress (1962) and Clack (1967) have reviewed the assumptions involved in a vector-summation model in some detail, those assumptions will not be reiterated here.

As we view the concepts of harmonic distortion and vector summation, they adequately predict a sinusoidal monaural phase effect similar to that exhibited in the results of the present experiments. The concepts of harmonic distortion in their simple form, however, together with a vector-summation model, do not predict, nor do they explain the frequency-dependent phase effects that were demonstrated in Experiment II, that is, the larger phase effects at low frequencies than at high frequencies. Vector summation predicts that summation between two signals of equal frequency will be independent of their frequency, and therefore, cannot explain frequency-dependent phase effects. Neither can quadratic nonlinearity nor the saturation type of nonlinearity, described earlier as peak clipping, explain frequency-dependent phase effects. Such nonlinearities, which can be described in terms of polynomial approximations, affect the input waveform in the same way, no matter at what frequency the signal may occur (Dallos, 1973). If we are to explain frequency-dependent monaural phase effects with traditional concepts of harmonic distortion and vector summation, we must then involve multivalued nonlinearities which, in general, cannot be approximated by a simple polynomial series and consequently are extremely complicated. Or, we must look for frequency dependencies not directly related to the locus of nonlinear distortion.

Maximum and Minimum Masking Phases

The vector-summation model predicts a cancellation phase of 180° and an augmentation phase of 0°. In the present experiments, maximum masking occurred with a cancellation phase of 90° ($\phi = \pi/2$) and minimum masking occurred with an augmentation phase of 270° ($\phi = 3\ \pi/2$). If we assume vector summation is a valid concept, then we must account for either a 90°-phase shift in the auditory system for the test signal relative to the masking signal, or we must account for a 90°-phase term in the second aural harmonic distortion product.

There are several apparent difficulties associated with the conclusion that maximum and minimum masking occur at 90° and 270°, respectively. First, it could be claimed that since the results of the present study were based upon few subjects, some question may remain whether the 90° and 270° phases for maximum and minimum masking apply to all listeners, or at least to most listeners. Both trained listeners in the present experiment showed maximum and minimum masking at 90° and 270°, respectively, as did four outpatient listeners who were tested for phase effects. In a similar experiment reported by Clack (1967), maximum and minimum masking phases from three listeners averaged 60° and 240° when expressed in phase terms consistent with the present study instead of in arbitrary phase terms. Using a slightly different technique, Kameoka and Kuriyagawa (1968) obtained maximum and minimum phases from two listeners that were around 85° and 265°. All three of those studies used signals that are comparable. The maximum and minimum masking phases those studies reported are closer to the 90° and 270° phases found in the present study than to the 0° and 180° phases required by a vector-summation model.

Further evidence that the 90° and 270° waveforms are somewhat unique to the hearing mechanism comes from data reported by Craig and Jeffress (1962). For signals comparable to those used in the studies mentioned above, with an f_1 signal at 60 dB SPL and an f_2 signal at low levels (13, 23, and 33 dB SPL), their two subjects could discriminate between the two signals of opposite phases much better when the two contrasting phases were 90° and 270° than when any other pairs of opposite phase conditions were presented. Since these investigations tend to support our findings of maximum and minimum masking phases at 90° and 270°, we are confident that our findings will stand the test of further experimentation.

Frequency-Dependent Phase Shift?

If we then accept the 90° and 270° estimates of maximum and minimum masking phases as those expected from most normal-hearing listeners, then we are left with the second problem of explaining the 90°-phase difference. It could be postulated that a 90°-phase shift exists in the auditory system that is a relative function of frequency, that is, a 90°-phase shift for the f_2 test signal relative to the f_1 masking signal. We also could postulate that the 90°-phase difference exists in the distortion process itself.

Evidence of a frequency-dependent phase shift of 90° per octave in the auditory pathway is meager. In fact, Nordmark, Glattke, and Schubert (1969) reported some convincing findings in guinea pig cochleas that argue against any such phase shift. They found that a 90° $f_1 + 2f_1$ waveform, such as that shown in Figure 2 of the present study, was essentially preserved in the electrical waveform measured across a pair of differential cochlear electrodes in both Turn I and Turn III of the cochlea. Only when the frequency of the $2f_1$ tone was considerably higher than the frequency that produced maximal

cochlear microphonic (CM) response from the Turn III electrodes did a frequency-dependent phase shift even begin to occur. Tonndorf (1970) and Dallos, Schoeny, and Cheatam (1971) have argued that the frequency-dependent phase shift that Nordmark et al. recorded was essentially a limitation of the CM recording. Most recent evidence (Dallos and Cheatam, 1971) has shown that different frequency components propagate along the basilar membrane with the same velocity, which effectively argues against a 90°-per-octave frequency-dependent phase shift for the signals concerned with here. We are left with the interpretation that whatever type of aural distortion is reflected in octave-masking experiments, it is the distortion product itself that accounts for this 90°-phase difference.

Traveling Waves and Aural Harmonic Distortion

Before considering alternative models that take frequency-dependent effects and the 90°-phase difference into account, we should note here that traditional theories of aural harmonic distortion imply that a traveling wave exists at the place on the basilar membrane corresponding to the frequency of that aural harmonic distortion component. Experiments by Wever, Bray, and Lawrence (1940a, b; 1941) have ruled out middle-ear mechanisms as a significant source of nonlinear distortion, except perhaps at extreme levels, so we cannot make a case for traveling waves generated by middle-ear distortion. Tonndorf (1969) reported observations of eddies in cochlear models which support the hypothesis that hydrodynamic coupling processes actually produce traveling waves in the cochlea at the region on the basilar membrane that corresponds to the frequencies of the harmonic distortion components. It would be attractive to postulate a frequency-dependent 90°-phase shift in a hydrodynamic coupling mechanism that produces a traveling wave at the place on the basilar membrane corresponding to the frequency of the distortion product.

Recent CM data, however, obtained by Dallos and Sweetman (1969), relegate hydrodynamic distortion processes to relatively high levels of stimulation (above 80-90 dB SPL) and provide convincing evidence that low-level nonlinear distortion products are maximal at the place on the basilar membrane most responsive to the fundamental tone, instead of at the place most responsive to the frequency of the harmonic distortion product. While presenting tones to the guinea pig cochlea at 60 dB SPL, Dallos and Sweetman recorded CM amplitude distribution functions of various CM harmonics. They found that the CM harmonics were most prominent at the location in the cochlea where the eliciting primary was located. They also found that CM distortion components could not be simultaneously cancelled at two cochlear locations, a finding which led them to conclude that harmonic distortion components do not generate traveling waves of their own.

Distortion in the Generator Potential

Dallos and Sweetman's (1969) results are consistent with the hypothesis

that places the major source of nonlinearity at the level of individual hair cells. Whitfield and Ross (1965) have suggested that the operating characteristic of each hair cell is nonlinear and symmetrical. The wave forms of the generator potentials produced by each hair cell, driven at moderate levels, can be described by a distorted sinusoid made up of a dc potential, a sinusoid of the basic frequency, and harmonics at higher frequencies. They further argue that the distorted waveforms of the individual generator potentials are not always seen in the CM because the electrodes average the ac components from a group of generators from neighboring regions of the basilar membrane. Because of propagation delay along the basilar membrane, those individual generator waveforms are not all in phase. Since the rate of change of phase is greater for high-frequency components, the higher harmonics from adjacent generators tend to cancel each other. The end result, as Whitfield and Ross (1965) suggest, is that the CM recording electrodes essentially look at the generator waveforms through a low-pass filter, and therefore what is recorded from a group of generators by a single pair of differential cochlear electrodes looks like an undistorted waveform.

Now, if each generator potential reflects a nonlinear transfer function, then it is reasonable to expect that the generators which receive the largest displacement from a given stimulus will also reflect the most harmonic distortion or waveform distortion. The place on the basilar membrane corresponding to the frequency of the masking signal or eliciting tone produces the largest traveling-wave displacement amplitude. Therefore, as Dallos and Sweetman (1969) have shown, maximum harmonic distortion should occur at that place on the basilar membrane corresponding to the frequency of the masking signal, and not at the place on the basilar membrane corresponding to the frequency of both the aural-distortion component and the test signal.

This should not imply that harmonic-distortion components are only localized to the region of the basilar membrane of maximum displacement. All generators stimulated by a traveling wave should exhibit the same nonlinearity relative to the magnitude of their displacement. Those generators stimulated by the basal portion of a traveling wave should not exhibit as much absolute distortion as those generators stimulated by the more apical peak of the traveling wave because the absolute displacement in the basal tail of the traveling wave is less. However, once the level of the stimulating tone becomes high enough, basal generators should be just as capable of producing distortion products as the more apical generators.

What emerges from this examination of recent physiological literature is the concept that each hair cell stimulated by a given masking signal is capable of nonlinear waveform distortion, and that the relative contribution of the distortion from an individual generator to the total harmonic distortion in the auditory system depends upon the relative magnitude of displacement of that generator. Maximum CM harmonic distortion is recorded at the place of maximal displacement (f_1 in the octave-masking paradigm), but this does not preclude similar distortion existing at f_2 (when $f_2 = 2f_1$). Since in octave

masking the f_2 test signal is considerably less intense than the f_1 masker, there is little chance for interaction between the displacement produced by the test signal at an f_2 place and any aural distortion component produced by the masking signal at an f_1 place on the basilar membrane. It is more likely that the interaction occurs between the displacement produced by the f_2 test signal at the f_2 place and the harmonic distortion components of the f_1 masking signal at the f_2 place. If Whitfield and Ross's (1965) individual generator theory of nonlinear distortion is true, then no traveling wave at the place on the basilar membrane corresponding to the distortion frequency is required.

Considering the frequency-dependent phase effects and the frequency-dependent masking slopes found in the present octave-masking data, and considering the recent physiological evidence arguing against the existence of a traveling wave at the frequency of an aural distortion product, we feel that some alternative explanations are in order.

Phase of Aural Distortion

Let us first attack the problem of accounting for a 90°-phase term in the second harmonic-distortion product, since we have already ruled out a 90°-frequency-dependent phase shift. A clue to the absolute phase of the second harmonic-distortion product in the hypothetically distorted generator potential described earlier can be found by considering the output waveform when a sinusoid is passed through an electrical device with a nonlinear (sigmoidal) transfer function, a device which exhibits saturation or amplitude limiting. In addition to the nonlinear input-output characteristic, assume that the operating point of that electrical device is also asymmetrical. The resulting output wave form from that device will appear peak clipped on one-half of the wave. Fourier's analysis of that waveform will show an f_1 fundamental component plus other harmonic components at $2f_1$, $4f_1$, and so on, at decreasing amplitudes. The phase of the $2f_1$ component, or second harmonic, will be 90° ($\pi/2$).

A more intuitive approach would be to substitute a hair cell for the electrical device described above. The half-wave peak-clipped distorted wave form of that hair cell generator would appear much like the electrical wave form that results when an f_1 sinusoid and a $2f_1$ sinusoid are added together with the $2f_1$ sinusoid less intense (by about 10 dB) and with a phase shift of 90° in the $2f_1$ sinusoid. The nonlinear transfer function of that generator can be attributed, for example, to the inability of the basilar membrane to follow large displacements in one direction, in other words, to displacement limiting. The point is that the output waveform from a distorted generator with an f_1 input appears much like the output waveform of an undistorted generator when the input signal is $f_1 + 2f_1$ with a phase shift of 90° between the two signals. Increased distortion would make the output waveform look more and more like the 90° $f_1 + 2f_1$ waveform shown in Figure 2.

This model of individual hair cell or generator distortion predicts a certain

phase relation or phase shift between a fundamental and its harmonics. For the $2f_1$ harmonic of an f_1 fundamental, the phase shift predicted is 90°, or $\pi/2$ radians, a prediction we feel accounts for the 90°-phase difference described in the octave-masking data reported above and which takes into account recent physiological data that argue against a traveling wave at the place on the basilar membrane corresponding to the frequency of the distortion product.

Neural Pattern Recognition and Frequency-Dependent Effects

At this point in our discussion, vector summation between the waveform of a $2f_1$ external test signal and the distorted f_1 output waveform of individual generators at the $2f_1$ place on the basilar membrane could easily account for the sinusoidal monaural phase effect. However, vector summation alone would not account for the frequency-dependent phase effects described earlier. If the discriminations between masking and test signals that are involved here were to take place at a higher stage in the auditory system, at a stage where phase relations among components might be transformed into temporal patterns, say in the firing patterns of neural discharges, then frequency-dependent monaural phase effects might be explained adequately in terms of the temporal resolving power of that part of the auditory system (temporal resolution being better at low than at high frequencies). This is not a new conjecture. Craig and Jeffress (1962) included "temporal pattern" analysis at a neural stage as one possible model to explain monaural phase effects. Raiford and Schubert (1971) also portrayed a similar temporal model which they refer to after Goldstein (1966) as "place intensity" analysis.

Since Craig and Jeffress (1962) interpreted their results, new neurophysiological data have become available which apply directly to this problem. Brugge et al. (1969) obtained data on temporal coding from auditory nerve fibers of squirrel monkeys using signals similar to those concerned with here, that is, f_1, $2f_1$, and $f_1 + 2f_1$ as a function of phase. Their phase-locked data from Unit 67-154-13 (taken from Brugge et al., Figure 1), plotted as folded period histograms in Figure 11, demonstrate conclusively that the phase relations in $f_1 + 2f_1$ signals are coded in the discharge of single auditory neurons as temporal patterns.

The histograms in Figure 11 are for a neural unit whose best frequency was midway between f_1 and $2f_1$. The f_1 signal (907 Hz) was at 50 dB SPL and the $2f_1$ signal (1814 Hz) was at 60 dB SPL (an $f_1/2f_1$ intensity ratio of -10 dB). The abscissas of the folded period histograms in Figure 11 have been shifted from those in the original plots to make the lowest spike rates in each histogram correspond with each other. Those shifts also serve to line up the negative-going zero crossings of the input waveforms. The signal conditions, of course, are not identical to those that existed in the octave-masking paradigms of the present study. Since neural data obtained with signal conditions parallel to those employed in octave masking do not exist, we must assume

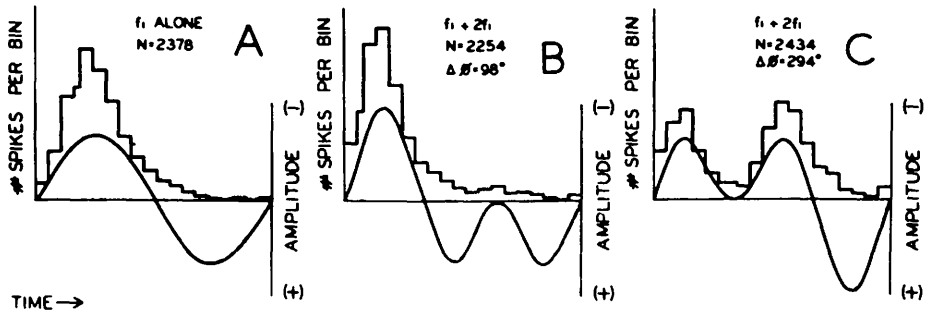

FIGURE 11. Period histograms of neural discharges from an auditory nerve fiber in squirrel monkey. Neural data are from Brugge, Anderson, Hind, and Rose (1969). One period of the complex wave form is represented on the abscissa, and the number of spikes that occurred during each time-bin is shown on the ordinate. "N" specifies the total number of spikes obtained during the 10-second presentation of each waveform. $\Delta\phi$ indicates the phase relations between the f_1 and $2f_1$ signals that were used to obtain the neural data for each histogram. Superimposed on each period histogram is the waveform corresponding to each of three signals employed in the octave-masking experiments: f_1 alone (A); 90° $f_1 + 2f_1$ (B); and 270° $f_1 + 2f_1$ (C). The abscissa in each histogram has been shifted from the original plots (Brugge et al., Figure 1) simply to make the negative going zero crossings of their computer-fit sinusoids correspond with each other. Positive amplitude corresponds to an inward deflection of the eardrum.

that neural results similar to those shown in Figure 11 would be obtained if $2f_1$ were closer to the best frequency of that neural unit, and if the signals employed were to have an $f_1/2f_1$ intensity ratio in the range of 20 to 50 dB, as in the octave-masking experiments. Examination of the response areas reported by Brugge et al. for Unit 67-154-13 indicates that the unit would have responded under such signal conditions, although with decreased temporal resolution.

Recall that in the 4IFC listening paradigm, used in the octave-masking experiments, the f_1 waveform was presented alone in three intervals and the $f_1 + 2f_1$ waveform was presented in the "odd" interval (see Figure 3). A listener simply had to decide which interval was different, in any way, from the other three. The waveform for f_1 alone and the corresponding single-unit discharge pattern expected from a neural unit with a best frequency midway between f_1 and $2f_1$ are shown in Figure 11A. The waveform for $f_1 + 2f_1$ with $\phi = 90°$ and the corresponding single-unit discharge pattern are shown in Figure 11B. Similarly, the displacement waveform for $f_1 + 2f_1$ with $\phi = 270°$ and the corresponding single-unit discharge pattern are shown in Figure 11C.

Notice that an asymmetry exists in the signal waveform, that is, the bottom half of the waveform below the abscissa is different from the top half. That asymmetry is unique only for a two-tone signal with a frequency ratio of 1:2 and is largest for the 90° and 270° phases. That asymmetry is also reflected in the neural discharge patterns of Figures 11B and 11C. Grossly different

temporal patterns occur for the two opposite phases. Because the auditory system apparently behaves like a half-wave rectifier at some point before the neural response, only one-half the waveform is reflected in the temporal discharge pattern. The most disparate temporal patterns occur for only one pair of opposite phases, 90° and 270° as shown in Figures 11A and B. All other pairs of opposite phases do not result in temporal firing patterns as different from one another as do 90° and 270°.

In descriptive terms, consistent with the behavior of neural discharge patterns, a listener's task during the two opposite phase conditions of the octave-masking experiments was essentially (1) to discriminate between the temporal pattern in Figure 11A and the temporal pattern in Figure 11B during the 90°-phase condition, and (2) to discriminate between the temporal pattern in Figure 11A and the temporal pattern in Figure 11C during the 270°-phase condition. A listener would be expected to have the most difficulty discriminating between temporal patterns that are most similar, in this case between the f_1 pattern in 11A and the 90° $f_1 + 2f_1$ pattern in 11B. Indeed, in the octave-masking experiments reported earlier, it was the 90° phase that was most difficult. The 90° conditions needed more intense $2f_1$ (f_2) signals than the 270° conditions before 50% of the discriminations between the f_1 and $f_1 + 2f_1$ waveforms could be made correctly. Similarly, waveforms with the least similar temporal patterns (11A vs 11C), corresponding to f_1 alone and to $f_1 + 2f_1$ with a 270° phase, were those which needed much less intense $2f_1$ signals for 50% correct discrimination. Incidentally, the 90° and 270° phases were also the two opposite phases which could be identified most easily in Craig and Jeffress' (1962) experiments.

Temporal pattern similarities, described as period histograms above, can be described in terms of the raw neural data as well, that is, as eighth-nerve interspike intervals. The interspike intervals from a unit at the $2f_1$ place on the basilar membrane average around $1/f_1$ sec for both the f_1 signal alone (Figure 11A) and for the 90° $f_1 + 2f_1$ signal (Figure 11B), which are highly indiscriminable intervals except perhaps for the dispersion around the $1/f_1$ sec mean. In contrast, the interspike intervals average around $1/2f_1$ sec for the 270° $f_1 + 2f_1$ signal (Figure 11C). These are highly discriminable intervals when paired with intervals of $1/f_1$ sec for the f_1 signal alone (Figure 11A).

It is well known that neural units follow the temporal characteristics of an amplitude waveform much more precisely for low-frequency waveforms than for high-frequency waveforms (Rose, 1970). Similarly, in periodicity-pitch experiments, pitch phenomena attributed to the periodicity in an amplitude waveform are more pronounced for low-frequency signals than for high-frequency signals (Small, 1970; Ritsma, 1970). It is not surprising, then, that the phase effects seen in the present study are larger at low frequencies than at high frequencies. The neural discharge patterns are simply better defined and, therefore, more dissimilar at lower frequencies because temporal resolution is better when longer interspike intervals are involved.

During the 90°-phase conditions, the f_1 temporal pattern (Figure 11A) and

the $f_1 + 2f_1$ temporal pattern (Figure 11B) are highly similar. The most obvious mechanism for discrimination between these two patterns may simply be the total output at some point in the system (energy detection), since little difference exists between the two patterns in the temporal domain. No obvious clues are evident in the temporal patterns of the 90°-phase conditions that would explain the frequency-dependent monaural phase effects reported here.

The temporal firing patterns during the 270°-phase conditions, however, do offer explanations of frequency dependencies in monaural phase effects. The f_1 temporal pattern (Figure 11A) and the 270° $f_1 + 2f_1$ temporal pattern (Figure 11C) are highly dissimilar in the temporal domain, as stated earlier. Because temporal resolution in the auditory system is much more accurate at low frequencies that at high frequencies, the 270° $f_1 + 2f_1$ pattern will be more disparate from the f_1 pattern at low frequencies than at high frequencies. Therefore, one would predict that at low frequencies, little $2f_1$ signal at 270° need be added to an f_1 signal to produce a discriminable pattern disparity. At high frequencies where temporal resolution is less accurate, a more intense $2f_1$ signal at 270° must be added to an f_1 signal to produce disparate temporal patterns. If discrimination between f_1 and $f_1 + 2f_1$ patterns in the 90°-phase condition is based upon energy detection, a relatively frequency-independent process, then the differences in signal intensity needed to produce discriminable pattern differences in the 270°-phase conditions at low vs high frequencies would account for the frequency dependent phase effects.

Neural Pattern Recognition and Phase Effects

Signal waveforms and neural-temporal patterns corresponding to maximum and minimum octave masking appear to agree, at least to the first approximation allowed by existing neural data. This agreement between the absolute phases that produce the most disparate neural patterns and the phases that produce maximum and minimum masking is convenient since it helps explain the 90°-phase difference mentioned earlier. However, only the disparate neural discharge patterns unique to only one pair of opposite phases are needed to explain the cyclical variations in discriminability (or masking) that occur as a function of phase in phase-locked octave-masking experiments. One must only assume that temporal pattern recognition varies on a continuum from most similar to least similar to account for the phase-dependent sinusoidal change in masked thresholds at $2f_1$.

Slope of Masking

Masking slopes for test frequencies near the masker and above the masker (upward spread of masking) have been documented by Wegel and Lane (1924), Fletcher (1930), Egan and Hake (1950), and Small (1959) for pure-tone maskers, and by Egan and Hake (1950), Bilger and Hirsh (1956),

and Jerger, Tillman, and Peterson (1960) for bands of noise. Masking grows with a slope near 1.0, that is, 1 dB of masking per 1-dB change in the level of the masker, when the two signals are close in frequency. Masking grows with slopes as high as 3.0 dB/dB when test signals are at higher frequencies than the masker, whether the masker is a band of noise or a pure tone. Results typical of those masking experiments are replicated in the mistuned masking data of Figure 8. The masking slope for the 1006 Hz test signal masked by a 1000-Hz pure tone is about 0.80 dB/dB, while the steepest slope for the mistuned octave test signal at 2006 Hz is 3.0 dB/dB. That steep masking slope occurs over a small range of f_1 masker levels, from about 65 to 75 dB SL. Note that the range of f_1 masker levels over which the steep masking occurs is relatively constant for test signals from 1503 through 3006 Hz. Note, also, that the level above the threshold at which the steep masking begins becomes greater as the frequency of the test signal moves further away (higher) from the masker frequency.

Previous explanations for these nonlinear masking functions have relied heavily upon nonlinearities that could be described with a polynomial series. Those polynomial nonlinearities generate harmonic-distortion components which produce masking that grows at a rate greater than 1 dB/dB, a rate dependent upon the order of the distortion component. Masking at the second harmonic should grow at the rate of 2 dB/dB and at the third harmonic at 3 dB/dB. If that explanation were true, one might expect steeper masking slopes at the frequencies of the harmonic-distortion components than at frequencies midway between those harmonic components where supposedly harmonic-distortion components are less intense. One might also expect steeper slopes at $3f_1$ than at $2f_1$ because the harmonic order of the distortion product is one higher. As shown in Figure 8, neither is the case. The masking slopes are about the same for test signals at 1503, 2006, 2503, and 3006 Hz. The only points on the different functions which change consistently with frequency of the test signal are those levels of the f_1 masker where steep masking begins and ends. Stated differently, the only difference except at high f_1 masker levels among the masking functions for f_2 = 1503, 2006, 2503, and 3006 Hz, is the level of the f_1 masker at which "aural distortion" begins to be detectable. There are no gross differences in the slopes of masking as the frequency of the test signal increases.

Consideration of traditional concepts of auditory theory yields an alternative explanation for the steep masking slopes. Bekesy's (1960) observations of cochlear models, and his as well as Rhode's (1971) more recent observations of basilar membrane vibrations, have established that the envelope of displacement on the basilar membrane is asymmetrically skewed toward the base. In other words, as the level of a signal is increased, new regions of the basilar membrane become involved at a much faster rate basalward than apically to the place on the membrane corresponding to maximal displacement. Critical-bandwidth experiments (Fletcher, 1940) have demonstrated that masking increases proportional to the bandwidth of the masking noise up to a

critical bandwidth, beyond which further increases in bandwidth do not produce more masking.

Consider now the situation that exists when an f_1 masker just begins to mask an f_2 test signal (where $f_2 > f_1$). At that point, the moderate-level (about 60-70 dB SL) f_1 displacement pattern probably just barely overlaps into the region on the basilar membrane displaced by the low-level f_2 displacement pattern (0-10 dB SL). Neither the f_2 test signal nor the f_1 masker is likely to involve a large enough area of the basilar membrane at the f_2 place to activate all of the neural units that may contribute to masking at higher levels. That is, they do not involve all the neural units that interact together to make up a critical masking band at the f_2 place.

As the level of the f_1 masker is increased, two things occur simultaneously that are likely to contribute to masking at f_2. First, the amplitude of the f_1 displacement pattern increases in the same area of the basilar membrane which contributed to masking of f_2 at the lower f_1 level. Second, the area of the f_1 displacement pattern increases with f_1 level, thereby involving new areas of the basilar membrane and new neural units that were not stimulated before at the lower f_1 level. Those new neural units, however, are still within the critical masking band at the f_2 place. We now have two masking components that work together to produce at least twice as much masking. One masking component can be called displacement amplitude, and the other can be called displacement area. Narrower displacement envelopes exist for higher frequencies and may account for a more significant contribution by the displacement-area masking component at higher frequencies and, therefore, produce steeper masking slopes with higher fundamental frequencies (see Figure 6).

The saturation of octave masking at high levels of f_1 might be explained as follows. After f_1 level is raised to about 70-80 dB SL, where the steep portions of the masking functions end, further increases in f_1 level probably involve only one of those masking components. At those high f_1 levels, the f_2 area on the basilar membrane that is displaced by the basal tail of the f_1 masker has probably become large enough to involve all the neural units that interact together to make up a critical masking band at the f_2 place. Further increases in the f_1 masker result in masking by only the displacement-amplitude masking component, since the new areas of the basilar membrane displaced by the basal tail of the f_1 masker are far enough removed from the f_2 place to preclude any interaction that contributes to f_2 masking. In simple terms, the critical masking band at f_2 has been exceeded and no further increases in masking bandwidth or displacement area on the basilar membrane will produce more masking of an f_2 signal.

In summary then, we have offered alternative explanations for octave-masking data that could not be adequately explained with classical concepts of auditory nonlinearity. The nonlinearities concerned with here have been localized to individual hair-cell generator potentials. The sinusoidal monaural phase effect has been attributed to temporal pattern discrimination. The 90°-phase difference in the octave-masking data has been attributed to the

nature of individual hair-cell waveform distortion and to the uniqueness of the neural discharge patterns corresponding to the 90° and 270° $f_1 + 2f_1$ waveforms. The frequency dependencies of the monaural phase effects in octave masking have been attributed to frequency dependence of temporal resolution in the neural system. Finally, the steep growth of masking at frequencies above the masker has been attributed to the asymmetrical nature of basilar membrane displacement patterns and to the fact that interaction among neighboring neural units on the basilar membrane is limited to what can be called a critical bandwidth or a critical masking band. We do not presume to label these explanations with a single concept or to lump them into a single model, since the explanations are still little more than speculative reasoning. However, we do propose that these explanations, which are modifications of common concepts among currently accepted auditory theories, do account for octave-masking data that are not as easily explained with the traditional models.

ACKNOWLEDGMENT

This investigation was supported by Grants NS 04105 and NS 07790 from the National Institute of Neurological Diseases and Stroke. The authors wish to thank Gerald Studebaker, Memphis State University, for his efforts in soliciting and evaluating independent criticism of this work. David A. Nelson's current address is Hearing Research Laboratory, Department of Otolaryngology, University of Minnesota, 2630 University Avenue, SE, Minneapolis, Minnesota 55414.

REFERENCES

BEASLEY, W., The monaural phase effect with pure binary harmonics. *J. acoust. Soc. Amer.*, 1, 385-402 (1930).

VON BEKESY, G., *Experiments in Hearing.* (E. G. Wever, Ed.) New York: McGraw-Hill (1960).

BILGER, R. C., and HIRSH, I. J., Masking of tones by bands of noise. *J. acoust. Soc. Amer.*, 28, 623-630 (1956).

BORING, E. G., *Sensation and Perception in the History of Experimental Psychology.* New York: Appleton-Century-Crofts (1942).

BRUGGE, J. F., ANDERSON, D. J., HIND, J. E., and ROSE, J. E., Time structure of discharges in single auditory nerve fibers of the squirrel monkey in response to complex periodic sounds. *J. Neurophysiol.*, 32, 386-401 (1969).

CHAPIN, E. K., and FIRESTONE, F. A., The influence of phase on tone quality and loudness; the interference of subjective harmonics. *J. acoust. Soc. Amer.*, 5, 173-180 (1934).

CLACK, T. D., Aural harmonics: The masking of a 2000-Hz tone by a sufficient 1000-Hz fundamental. *J. acoust. Soc. Amer.*, 42, 751-758 (1967).

CLACK, T. D., Aural harmonics: Preliminary time-intensity relationships using the tone-on-tone masking technique. *J. acoust. Soc. Amer.*, 43, 283-288 (1968).

CLACK, T. D., and BESS, F. H., Aural harmonics: The tone-on-tone masking vs. the best beat method in normal and abnormal listeners. *Acta Otolaryng.*, 67, 399-412 (1969).

CRAIG, J. H., and JEFFRESS, L. A., Effect of phase on the quality of a two-component tone. *J. acoust. Soc. Amer.*, 34, 1752-1760 (1962).

DALLOS, P., *The Auditory Periphery: Biophysics and Physiology.* New York: Academic (1973).

DALLOS, P., and CHEATHAM, M. A., Travel time in the cochlea and its determination from cochlear-microphonic data. *J. acoust. Soc. Amer.*, 49, 1140-1143 (1971).

DALLOS, P., SCHOENY, Z. G., and CHEATAM, M. A., On the limitations of cochlear microphonic measurements. *J. acoust. Soc. Amer.*, 49, 1144-1154 (1971).

DALLOS, P., and SWEETMAN, R. H., Distribution patterns of cochlear harmonics. *J. acoust. Soc. Amer.*, **45**, 37-46 (1969).
EGAN, J. P., and HAKE, H. W., On the masking pattern of a simple auditory stimulus. *J. acoust. Soc. Amer.*, **22**, 622-630 (1950).
FLETCHER, H., *Speech and Hearing*. New York: D. VanNostrand (1929).
FLETCHER, H., A space-time pattern theory of hearing. *J. acoust. Soc. Amer.*, **1**, 313-316 (1930).
FLETCHER, H., Auditory patterns. *Rev. mod. Physics*, **12**, 47-65 (1940).
FRICKE, J. E., Monaural phase effects in auditory signal detection. *J. acoust. Soc. Amer.*, **43**, 439-443 (1968).
GOLDSTEIN, J. L., An investigation of monaural phase perception. Doctoral thesis, Univ. of Rochester (1966).
JERGER, J. F., TILLMAN, T. W., and PETERSON, J. L., Masking by octave bands of noise in normal and impaired ears. *J. acoust. Soc. Amer.*, **32**, 385-390 (1960).
KAMEOKA, A., and KURIYAGAWA, M., P.S.E. tracing method for subjective harmonics measurement and monaural phase effect on timbre. *J. acoust. Soc. Amer.*, **39**, 1263A (1966).
NELSON, D. A., and BILGER, R. C., Pure-tone octave masking in listeners with sensorineural hearing loss. *J. Speech Hearing Res.*, **17**, 252-278 (1974).
NEWMAN, E. B., STEVENS, S. S., and DAVIS, H., Factors in the production of aural harmonics and combination tones. *J. acoust. Soc. Amer.*, **9**, 107-118 (1937).
NIXON, J. C., RAIFORD, C. A., and SCHUBERT, E. D., Technique for investigating monaural phase effects. *J. acoust. Soc. Amer.*, **48**, 554-556 (1970).
NORDMARK, J. O., GLATTKE, T. J., and SCHUBERT, E. D., Waveform preservation in the cochlea. *J. acoust. Soc. Amer.*, **46**, 1587-1588 (1969).
RAIFORD, C. A., and SCHUBERT, E. D., Recognition of phase changes in octave complexes. *J. acoust. Soc. Amer.*, **50**, 559-567 (1971).
RHODE, W. S., Observations of the vibration of the basilar membrane in squirrel monkeys using the Mossbauer technique. *J. acoust. Soc. Amer.*, **49**, 1218-1232 (1971).
RITSMA, R. J., Periodicity detection. In R. Plomp and G. F. Smoorenburg (Eds.), *Frequency Analysis and Periodicity Detection in Hearing*. Leiden, The Netherlands: A. W. Sijthoff (1970).
ROSE, J. E., Discharges of single fibers in the mammalian auditory nerve. In R. Plomp and G. F. Smoorenburg (Eds.), *Frequency Analysis and Periodicity Detection in Hearing*. Leiden, The Netherlands: A. W. Sijthoff (1970).
SMALL, A. M., Pure-tone masking. *J. acoust. Soc. Amer.*, **31**, 1619-1625 (1959).
SMALL, A. M., Periodicity Pitch. In J. V. Tobias (Ed.), *Foundations of Modern Auditory Theory*. Vol. I. New York: Academic (1970).
TONNDORF, J., Non-linearity in cochlear hydrodynamics. *J. acoust. Soc. Amer.*, **47**, 579-591 (1969).
TONNDORF, J., Comments on waveform preservation in the cochlea. *J. acoust. Soc. Amer.*, **48**, 596 (1970).
WEGEL, R. L., and LANE, C. E., The auditory masking of one pure tone by another and its probable relation to the dynamics of the inner ear. *Phys. Rev.*, **23**, 266-285 (1924).
WEVER, E. G., BRAY, C. W., and LAWRENCE, M., The locus of distortion in the ear. *J. acoust. Soc. Amer.*, **11**, 427 (1940a).
WEVER, E. G., BRAY, C. W., and LAWRENCE, M., The origin of combination tones. *J. exp. Psychol.*, **27**, 217-226 (1940b).
WEVER, E. G., BRAY, C. W., and LAWRENCE, M., The effect of middle ear pressure upon distortion. *J. acoust. Soc. Amer.*, **13**, 182-187 (1941).
WHITFIELD, I. C., and ROSS, H. F., Cochlear-microphonic and summating potentials and the outputs of individual hair-cell generators. *J. acoust. Soc. Amer.*, **38**, 126-131 (1965).
WOLF, R. V., Digitally programmable attenuator. *Behav. Res. Meth. Instr.*, **4**, 278 (1972).

Received August 1, 1972.
Accepted January 5, 1974.