Level-dependent critical bandwidth for phase discrimination

David A. Nelson
Department of Otolaryngology and Department of Communication Disorders, University of Minnesota, Minneapolis, Minnesota 55455

(Received 22 March 1992; accepted for publication 29 November 1993)

Monaural phase discrimination was evaluated at 1000 Hz in six normal-hearing listeners as a function of the frequency difference between components in three-tone complexes at 40, 60, 80, and 100 dB SPL. The phase of the center component of 100% sinusoidally amplitude-modulated (SAM) waveforms was shifted by 90° to produce quasi-frequency-modulated (QFM) waveforms that had identical long-term power spectra to the SAM waveforms but with different amplitude envelopes and temporal fine structure. At low modulation frequencies, where spectral components were close together and presumably all well within a single auditory filter, normal-hearing listeners could easily discriminate QFM from SAM waveforms. As modulation frequency increased, a point was reached where listeners could no longer distinguish QFM from SAM waveforms, referred to here as the critical bandwidth for phase discrimination (CB_{phs}). Discrimination performance (d') was measured as a function of modulation frequency to yield a psychometric function for phase discrimination. From that psychometric function, CB_{phs} was defined as the modulation frequency corresponding to $d'=1.0$. At a carrier frequency of 1000 Hz, CB_{phs} increased with level between 40 and 80 dB SPL according to the relation: $CB_{phs}=34\cdot I^{(0.136)}$. Above 80 dB SPL, very little change in CB_{phs} with level was seen. The increase with level of the geometric mean CB_{phs} was predicted from level-dependent auditory filter slopes inferred from forward-masked tuning curves, as was the tendency to reach an asymptote above 80 dB SPL. Comparisons with previous work indicate that CB_{phs} values from the best performing subjects were well within the audibility region for cubic difference tones. It is proposed that internally generated cubic difference tones interact with externally generated acoustic components, both limited by a level-dependent auditory filter, to produce an internal excitation envelope that is the basis for discriminating between SAM and QFM waveforms. It is also suggested that individual differences at low levels may be due to internal phase ambiguities.

PACS numbers: 43.66.Nm, 43.66.Mk, 43.66.Sr [WDW]

INTRODUCTION

Mathes and Miller (1947) demonstrated experimentally that the perception of monaural phase effects was associated with the limited frequency resolving power of the auditory system. They shifted the phase of the center component of a 100% sinusoidally amplitude-modulated (SAM) waveform by 90° to produce a quasi-frequency-modulated (QFM) waveform with the same long-term spectrum but with a different amplitude envelope and temporal fine structure. The phase changes in those three-tone complexes were heard as quality changes (differences in roughness or harshness) as long as the frequency separation between components did not exceed about 40% of the center component in the three-tone complex. For modulation frequencies exceeding about 40% of the carrier frequency, SAM and QFM waveforms were not discernible. That held true for center frequencies between 500 and 2000 Hz at a listening level of 60 dB SL. At lower listening levels, the critical modulation frequency at which SAM and QFM became indistinguishable was reduced. It was about 25% and 15% of carrier frequency at 40 and 20 dB SL, respectively. Goldstein (1967a) replicated the three-tone phase discrimination experiment with similar results. The critical modulation frequency for distinguishing QFM from SAM waveforms was shown to be proportional to carrier frequency and dependent upon the level of the complexes from 20 to 60 dB SL.

Because the long-term acoustic (external) spectra of SAM and QFM waveforms are identical, the most obvious clue to distinguishing these two types of waveforms was thought to be contained in the amplitude envelope or temporal fine structure realized by the phase shift. Goldstein (1967a) proposed an auditory filter model followed by envelope detection to account for his results and those of Mathes and Miller (1947). He suggested that all three components must interact sufficiently within the same auditory analyzing filter for the phase shift to produce a discernible difference in envelope or time structure. If one or more of the three components were attenuated sufficiently by the auditory filter, then the phase shift could not result in a different temporal structure. As modulation frequency increases, the frequency separation between components in SAM and QFM waveforms exceeds the auditory filter width at the carrier frequency, individual components begin to be separated into separate auditory filters, interactions among the three spectral components are minimized, and the temporal cues in the excitatory waveforms that lead to discriminations between SAM and QFM stimuli are reduced.

 0001-4966/94/95(3)/1514/11/$6.00

Since then, several investigations have implicated combination tones as important contributors to discriminations between SAM and QFM waveforms (Buunen and Bilsen, 1974; Buunen *et al.*, 1974; Buunen, 1975; Buunen *et al.*, 1977). Buunen has suggested that the discrimination between SAM and QFM waveforms is based upon changes in the internal spectrum, which is the result of interactions between internally generated combination tones and the externally generated acoustic components.

More recently, such phase discrimination experiments have been used to investigate temporal processing in listeners with cochlear hearing loss (Rosen, 1984, 1986, 1987; Rosen and Fourcin, 1986; Rosen and Smith, 1988). Although data were presented for only a few hearing-impaired subjects, results showed better phase discrimination performance from hearing-impaired listeners than from normal-hearing listeners, implying better temporal resolution by the hearing impaired. Similar findings were also reported for a few subjects on a slightly different phase discrimination experiment (Schroder and Leek, 1989). In general, these reports of better phase discrimination than normal are consistent with the notion that listeners with hearing losses demonstrate abnormal frequency selectivity in the form of broader than normal auditory filters, which allow interactions between components over a broader range of frequencies.

Potentially, phase discrimination experiments offer a way to assess both temporal resolution and frequency resolution in listeners with impaired peripheral auditory systems. However, the preliminary reports of level dependence (Mathes and Miller, 1947; Goldstein, 1967a), the potential contributions of combination tones (Buunen, 1975), and the nonlinear nature of normal frequency selectivity (Nelson, 1991), suggest caution in reaching conclusions about their performance without further investigation. In the present experiment, phase discrimination of SAM and QFM waveforms was measured from normal-hearing subjects as a function of modulation frequency over the range of stimulus levels necessary for testing hearing-impaired listeners (from 40 to 100 dB SPL). Critical bandwidths for phase discrimination were determined from the phase discrimination functions. Those critical bandwidths were then compared with model predictions based upon auditory filter parameters obtained with forward masking and with limits for combination-tone perception taken from previous studies.

I. METHOD

A. Stimuli

Examples of SAM and QFM waveforms used in the experiment are shown in Fig. 1. Pairs of SAM and QFM waveforms are shown for increasing modulation frequency from top to bottom. The top waveform of each pair shows the SAM stimulus. The bottom waveform shows the QFM stimulus. For the present investigation, SAM and QFM waveforms were constructed by summing sinusoids, which are mathematically described by Eq. (1):

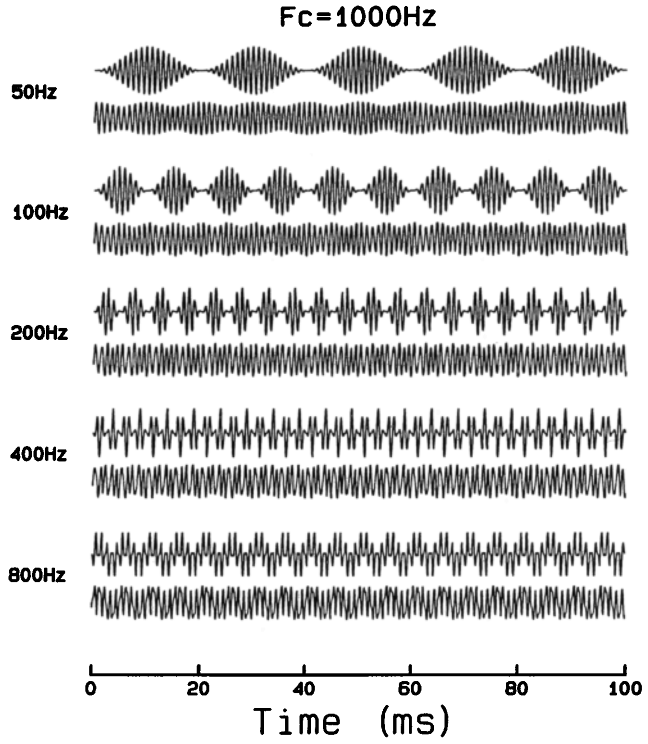

FIG. 1. Pairs of SAM and QFM waveforms at different modulation frequencies, as indicated to the left of each pair of waveforms. At each modulation frequency, the top waveform is the SAM stimulus and the bottom waveform is the QFM stimulus. Waveforms were calculated from Eq. (1). The QFM stimuli were produced by shifting the phase of the center component of the SAM stimuli by 90°. The long-term power spectra of each pair are identical.

$$s(t)=\alpha[0.5\sin(\omega_c-\omega_m)t-\sin(\omega_c t+\theta)+0.5\sin(\omega_c+\omega_m)t]. \quad (1)$$

The phase (θ) of the center frequency component (ω_c) is 0° for SAM and 90° for QFM. The carrier frequency is given by ω_c, the modulation frequency by ω_m, and the overall amplitude is given by α.

The long-term power spectra of SAM and QFM waveforms are identical, consisting of three frequency components, with upper (F_u) and lower (F_l) components equally spaced about the carrier frequency or center component (F_c) by an amount equal to the modulation frequency (F_m) and with an amplitude that is half the amplitude of the center component (-6 dB). With $\theta=0$, this corresponds to a 100% amplitude modulated waveform as illustrated in Fig. 1. The sign of the center component was negative to shift the envelope by 90° so that the SAM waveform began at zero. Waveforms were constructed mathematically in a PDP-8/E minicomputer, converted to analog waveforms with a 12-bit digital-to-analog converter at 50 μs per point, and then low-pass filtered at 8 kHz. Each stimulus was presented for 500 ms, including 10-ms rise and decay times. The overall levels of the SAM and QFM waveforms were presented at fixed sound pressure levels, as measured by a sound level meter in a 6-cc coupler. Stimulus levels were controlled by programmable attenuators.

Acoustic stimuli were presented monaurally through a TDH-39 earphone in an MX41/AR cushion. Subjects listened in a double-wall sound treated room.

Although SAM and QFM stimuli have identical long-term power spectra, their amplitude envelopes and their temporal fine structure differ. The SAM waveforms are characterized by sinusoidal modulations in envelope amplitude, which go through zero every $1/F_m$ ms. The QFM waveforms are characterized by very minor fluctuations in envelope amplitude, which reach a minimum every $1/(2F_m)$ ms.

B. Procedure

A 4AFC method of constant stimuli was employed to determine percent correct discrimination between SAM and QFM waveforms as a function of modulation frequency. Subjects listened to sounds in four sequential intervals indicated by lights. Three of the intervals contained the SAM signal and one of the intervals, determined randomly from trial to trial, contained the QFM signal. Subjects indicated which interval contained the "different" signal by pressing one of four buttons, after which they received correct-answer feedback and a new trial began. Time between signal intervals within a trial was 250 ms. During a single listening session, a complete psychometric function was obtained for phase discrimination as a function of modulation frequency. A single session consisted of 20 trials per modulation frequency, which were presented in ascending order from low- to high-modulation frequency. Two sessions were conducted per presentation level to yield 40 trials per condition for the three subjects who were tested in both ears at 40, 60, 80, and 100 dB SPL. Four sessions were conducted to yield 80 trials per condition for an additional three subjects who were tested in one ear at 20-dB sensation level (SL) as well as at 40, 60, 80, and 100 dB SPL. All subjects received at least one session of practice at each level before data were collected for the experiment.

II. RESULTS

A. Phase discrimination functions from individual ears

Figure 2 shows the results from the left ear of subject DJ for a carrier frequency of 1000 Hz. Percent correct discrimination between SAM and QFM stimuli is shown on the ordinate as a function of modulation frequency. The results at different stimulus levels are shown in different panels. Within each panel, the phase discrimination functions for two 20-trial-per-point tests are shown to illustrate how consistent the discrimination between SAM and QFM stimuli can be for individual ears. The general form of the phase discrimination function is well described by the results for DJ(L) at 100 dB SPL in Fig. 2. At low to moderate modulation frequencies, SAM and QFM stimuli are easily distinguished, with discrimination performance at 100%. As modulation frequency exceeds some critical value (around 400 Hz at 100 dB SPL or 40% of F_c) SAM

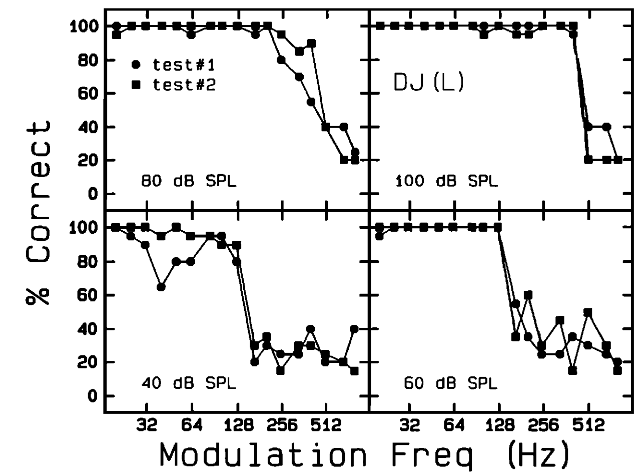

FIG. 2. Phase discrimination functions from the left ear of subject DJ. Percent correct discriminations between SAM and QFM stimuli are shown as a function of modulation frequency (frequency separation between components). Two separate tests are shown to demonstrate the repeatability of the discrimination task. Chance performance is 25%. The modulation frequency at which the break point occurs, between perfect performance and chance performance, increases with stimulus level.

and QFM stimuli become indistinguishable and discrimination performance drops to chance (25%).

At lower stimulus levels the perceptual differences between SAM and QFM stimuli become less robust. At low-modulation frequencies discrimination performance still remains at 100% over a range of modulation frequencies, but the break point between perfect performance and chance performance moves to lower modulation frequencies and the transition region becomes less abrupt. For example, at 40 dB SPL in Fig. 2 the break point occurs at about 128 Hz for DJ(L). Some evidence of learning from test to test is also seen for modulation frequencies around 40 Hz, since performance on the second test improved considerably over that shown on the first test.

The two sets of phase discrimination functions obtained from the left ear of subject KS are shown in Fig. 3. These functions have the same general form seen earlier for

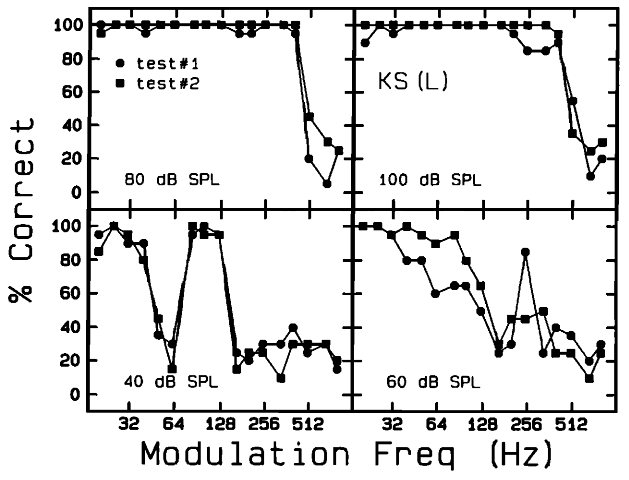

FIG. 3. Phase discrimination functions from the left ear of subject KS. A nonmonotonic function was evident at 40 dB SPL.

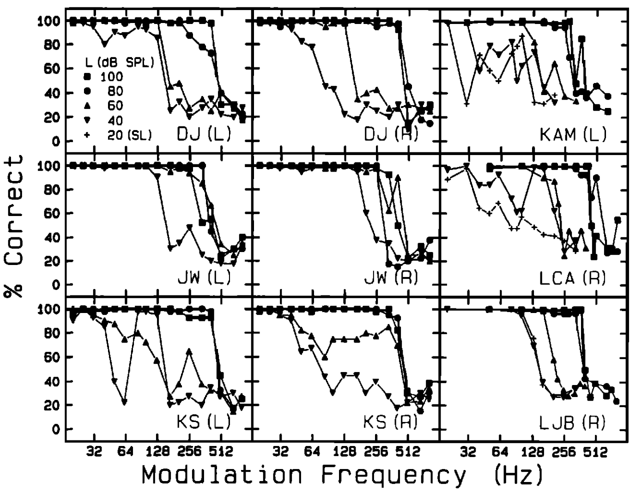

FIG. 4. Phase discrimination functions from both ears of the three normal-hearing subjects who were tested at 40, 60, 80, and 100 dB SPL are shown in the first two columns of panels. The third column of panels shows the phase discrimination functions from one ear of the three normal-hearing subjects who were tested at 20 dB SL and at 40, 60, 80, and 100 dB SPL.

DJ(L). The break point between perfect performance and chance performance moved toward lower modulation frequencies as stimulus level decreased. Evidence of learning from the first test to the second is also evident for this subject at moderate modulation frequencies (around 60 Hz) at 60 dB SPL. Of particular interest is the nonmonotonic function obtained from KS(L) at 40 dB SPL. For modulation frequencies below about 40 Hz, discrimination performance between SAM and QFM stimuli was excellent. At 50 and 62 Hz, performance dropped to chance and then increased abruptly to 100% again as modulation frequency increased to 83 Hz. Then, for modulation frequencies above 128 Hz, performance fell to chance again. As indicated by the close correspondence between the performance obtained for the two separate tests, this nonmonotonicity was consistent. Of the nine ears tested, nonmonotonic functions were evident at 40 dB SPL in three ears (KSL, KAM, and LCA).

To help quantify the break points in individual phase discrimination functions, performance scores on the two separate tests were combined to yield phase discrimination functions based upon 40 trials per point. These phase discrimination functions are shown in the first and second column of panels in Fig. 4 for the three subjects tested in both ears. It can be seen that combining scores across the two tests resulted in phase discrimination functions with the same general characteristics, i.e., the break point between distinguishable and indistinguishable stimuli was still abrupt. Also evident, is the tendency for results from one subject to be similar for both ears. In particular, even though the nonmonotonicity at 40 dB SPL for the left ear of KS is not exactly repeated in the right ear, the performance in the right ear is somewhat reduced over a large range of modulation frequencies compared to the performance seen in the other two subjects over the same range of modulation frequencies.

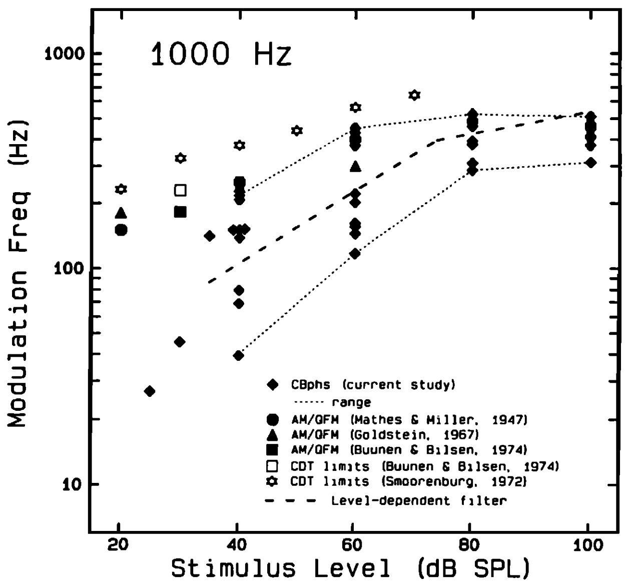

FIG. 5. Critical bandwidths for phase discrimination (CB_{phs}) as a function of stimulus level from the present study are shown by filled diamonds. The range of performance is indicated by dotted lines. Results from previous studies of SAM and QFM discrimination are shown by filled circles, filled triangles and a filled square. Limits for the perception of cubic difference tones are indicated by unfilled stars and an unfilled square. The dashed line shows predictions obtained with a level-dependent filter model using Eq. (6).

The phase discrimination functions from the three subjects tested only in one ear are shown in the third column of panels in Fig. 4. Notice that two other subjects (KAM and LCA) showed nonmonotonicities at 40 dB SPL. At 20 dB SL the functions became very gradual, with less well defined break points between perfect performance and chance, indicating that the task became quite difficult at that very soft level.

B. Critical bandwidth for phase discrimination (CB_{phs})

To specify the break points on the phase discrimination functions with some degree of objectivity and precision, performance scores were converted to d' values as a function of the logarithm of modulation frequency, and then linear least-squares fits were performed on the data within the transition region between perfect performance and chance performance. Those linear fits provided estimates of the modulation frequency corresponding to a d' of 1.0, which are referred to here as the critical bandwidth for phase discrimination (CB_{phs}). At modulation frequencies less than the CB_{phs}, SAM, and QFM stimuli are discriminable. At modulation frequencies greater than the CB_{phs}, SAM, and QFM stimuli are not discriminable.

The CB_{phs} values obtained from the nine normal hearing ears are shown by the filled diamonds in Fig. 5. Modulation frequency is shown on the ordinate as a function of the overall stimulus level in dB SPL. Table I lists the individual CB_{phs} values, their geometric means, and the modulation frequencies corresponding to one standard deviation from the geometric mean. As indicated by the filled diamonds in Fig. 5, and by the geometric means given in

TABLE I. Estimates of critical bandwidth for phase discrimination (CB_{phs} in Hz) based upon performance at $d'=1.0$. Results are from nine normal-hearing ears that were tested in quiet using AM and QFM complexes with the carrier frequency at 1000 Hz. Overall levels of the three-tone complexes are given in dB SPL, except at 20 dB sensation level (SL) where the levels in dB SPL are in parentheses.

	Level of SAM or QFM stimulus				
Subject	20 dB	40 dB	60 dB	80 dB	100 dB
DJL	···	150	162	458	471
DJR	···	79	156	525	448
JWL	···	150	430	394	375
JWR	···	219	450	308	407
KSL	···	152	117	475	508
KSR	···	69	373	487	455
KAM	27(25)	39	145	285	311
LCA	46(30)	209	222	495	466
LJB	141(35)	138	203	377	414
Geometric mean:	56	118	224	414	424
+1 s.d.	130	209	370	515	492
−1 s.d.	24	67	135	333	366

Table I, CB_{phs} increased with stimulus level for stimulus levels between 40 and 80 dB SPL. An asymptote was reached by 80 dB SPL, since the CB_{phs} values at 80 and 100 dB SPL were about the same. The range of values obtained from individual ears also decreased with level, as seen by the dotted lines in Fig. 5. The larger range of CB_{phs} values at 40 and 60 dB SPL is also evident in Fig. 4. For the purposes of description, the change in CB_{phs} seen in Fig. 5 between 40 and 80 dB SPL is well approximated ($R^2=0.9999$) by Eq. (2),

$$CB_{phs}=k\cdot I^m, \tag{2}$$

where I is the overall intensity of the three-component complex, $k=34.0$, and $m=0.136$.

III. DISCUSSION

These results indicate that SAM and QFM waveforms are easily discriminated at low and moderate modulation frequencies and become indistinguishable above some critical modulation frequency. That critical modulation frequency, referred to here as the critical bandwidth for phase discrimination (CB_{phs}), was shown to be strongly dependent upon the level at which SAM and QFM waveforms were presented. As indicated in Fig. 5, CB_{phs} increased with level between 40 and 80 dB SPL. These results confirm the findings of other investigators who obtained estimates of the modulation frequency at which SAM and QFM are no longer distinguishable (Mathes and Miller, 1947; Goldstein, 1967a; Nelson, 1977). The filled circles and filled triangles in Fig. 5 show the results at 1000 Hz reported by Mathes and Miller (1947) and Goldstein (1967a), respectively. They presented their stimuli at 20-, 40-, and 60-dB sensation level (SL). For the purposes of comparison in terms of dB SPL, we have assumed thresholds for detectability of the three-tone complexes near 1000 Hz in those two studies to be about 0 dB SPL. As seen in Fig. 5, their data reflect the same increase in critical modulation frequency with increases in stimulus level evidenced by the best subjects in the present study, up to 60 dB. Since their results only covered a range of stimulus levels from about 20 to 60 dB SPL, the present results extend the earlier findings to higher levels where we see a tendency for changes in CB_{phs} with level to reach an asymptote.

This CB_{phs} measure represents the frequency limits over which acoustic components interact sufficiently to allow discriminations between SAM and QFM waveforms. However, the mechanism or mechanisms underlying this discrimination are not so clear. There have been basically two general mechanisms proposed to account for these phase discriminations. Both involve modifications to the acoustic waveform by nonlinear auditory processing. One mechanism for discrimination is thought to be the temporal *excitation envelope* resulting from the interactions of the three acoustic components once they have been acted upon by a nonlinear (level-dependent) auditory filter. The other mechanism is thought to be the *internal spectrum* resulting from interactions between internal combination tones and external acoustic tones. Both mechanisms are considered below.

A. Predicting CB_{phs} from psychophysical tuning curves

The excitation envelope, as defined here, is the envelope of the excitation waveform after the individual components have been modified by the attenuation characteristics of the auditory filter. If the excitation envelope is indeed the basis for discrimination between SAM and QFM stimuli, then the changes in CB_{phs} with level seen here should be predictable from changes with level in characteristics of the auditory filter. To examine that premise quantitatively we evaluated a psychophysical model proposed by Goldstein (1967a), which predicts the critical modulation frequency from attenuation characteristics of the auditory analyzing filter.

Goldstein's (1967a) filter model involved an initial asymmetrical triangular filter, an ideal envelope detector to extract the excitation envelope, and a detection threshold based upon the power ratio of envelope maxima in SAM and QFM stimuli. In his model, the detection process required the determination of a "critical place" along the basilar membrane where the maximum envelope ratio between SAM and QFM would be the largest. For the SAM and QFM stimuli used in the present experiments, this always occurs when the upper and lower sidebands are equally attenuated relative to the center component. Because the auditory filter is assumed to be asymmetrical, the frequency corresponding to that critical place (F_{cp}) depends upon both the steep high-frequency slope and the more gradual low-frequency slope of the auditory filter. This relationship is shown in Eq. (3), which is adapted from Goldstein's equation (10), but for a triangular filter.

$$F_{cp}=F_c+(F_m*R_s), \tag{3a}$$

$$R_s=\frac{(S_2-S_1)}{(S_2+S_1)}, \tag{3b}$$

where F_c is the carrier frequency and F_m is the modulation frequency for SAM and QFM stimuli, and S_2 and S_1 are the high-frequency and low-frequency slopes of the auditory analyzing filter (in dB per Hz). The best frequency for discriminating SAM and QFM stimuli occurs just above the 1000-Hz center component, so the auditory filter responsible for the detection task must be centered just above the carrier frequency. Given the frequency of that critical place (F_{cp}), the relative attenuation (A) in decibels produced by the auditory filter can be determined for each of the three components, as given in Eq. (4):

$$A_l=S_1(F_{cp}-F_l), \tag{4a}$$

$$A_c=S_1(F_{cp}-F_c), \tag{4b}$$

$$A_u=S_2(F_u-F_{cp}), \tag{4c}$$

where F_l, F_c, and F_u are the frequencies of the lower, center and upper components of the three-tone SAM or QFM complex.

Calculation of the excitation envelope available to the auditory system after passing through the auditory filter near 1000 Hz requires some estimates of the slopes of the auditory filter. Goldstein evaluated his model using general estimates of those slopes from simultaneous masking patterns available to him in the literature at the time. S_2 was a constant 30 dB per critical band (CB) at all levels and S_1 was 20, 15, and 10 dB/CB at 20, 40, and 60 dB SL, respectively, with a critical band at 1000 Hz of 200 Hz. Because only the low-frequency slope was allowed to decrease with level, Goldstein concluded that the level dependence of the critical modulation frequency was attributable to changes in the low-frequency slope of the auditory filter with level.

More recent estimates of filter slopes at 1000 Hz can be inferred from psychophysical tuning curves obtained under forward masking (Nelson and Freyman, 1984; Nelson *et al.*, 1990; Nelson, 1991), which avoids some of the problems associated with simultaneous masking. Those estimates suggest that a triangular filter shape with two

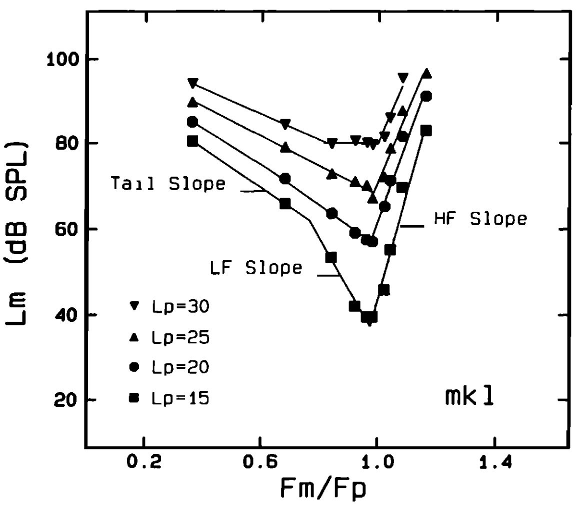

FIG. 6. Forward-masked psychophysical tuning curves reported by Nelson (1991). Level of the masker (L_m) is indicated on the ordinate as a function of the ratio between masker frequency (F_m) and probe frequency (F_p) on the abscissa. Probe frequency is 1000 Hz. The tuning curves were fit with three straight-line functions (in dB/Hz), one each for the tail slope, low-frequency (LF) slope, and high-frequency (HF) slope.

slopes (in dB per Hz) is a reasonable approximation at high stimulus levels, but at low levels a three-sloped function is more representative. This is evident in Fig. 6, which shows a series of forward-masked tuning curves from the left ear of a normal-hearing listener.[1] At each masker frequency in the tuning curve, the masker level is adjusted to produce a constant amount of masking at the probe frequency. Therefore, one can infer that the difference between the masker level at any given masker frequency and the masker level near the tip of the tuning curve describes the amount of attenuation provided by the auditory filter at that particular masker frequency. With these assumptions, and other assumptions about the absence of off-frequency listening (Nelson *et al.*, 1990), forward-masked tuning-curve slopes can be used to infer auditory filter slopes. The low- and high-frequency slopes of the tuning curve reflect the low- and high-frequency slopes of the auditory filter, respectively.

As shown in Fig. 6, at low operating levels where the masker levels near the tip of the tuning curve are around 40 dB SPL, the tuning curve is best described by a tail slope (S_{tl}), a low-frequency slope (S_{lf}), and a high-frequency slope (S_{hf}). At moderate levels and above, the low-frequency slope tends to merge with the tail slope, so the tuning curve is well described by a tail slope and a high-frequency slope. Nelson (1991) has provided quantitative estimates of these tuning-curve slopes over a wide range of operating levels from a group of normal-hearing listeners. The data reported in his Fig. 9 have been fitted with least squares procedures resulting in Eqs. (5a), (5b), and (5c), which specify changes in the three tuning-curve slopes as a function of the masker level near the tip of the tuning curve (L_{tip}):

$$S_{tl}=0.0820e^{(-0.0120L_{tip})}, \tag{5a}$$

$$S_{lf}=0.6089e^{(-0.0391L_{tip})}, \tag{5b}$$

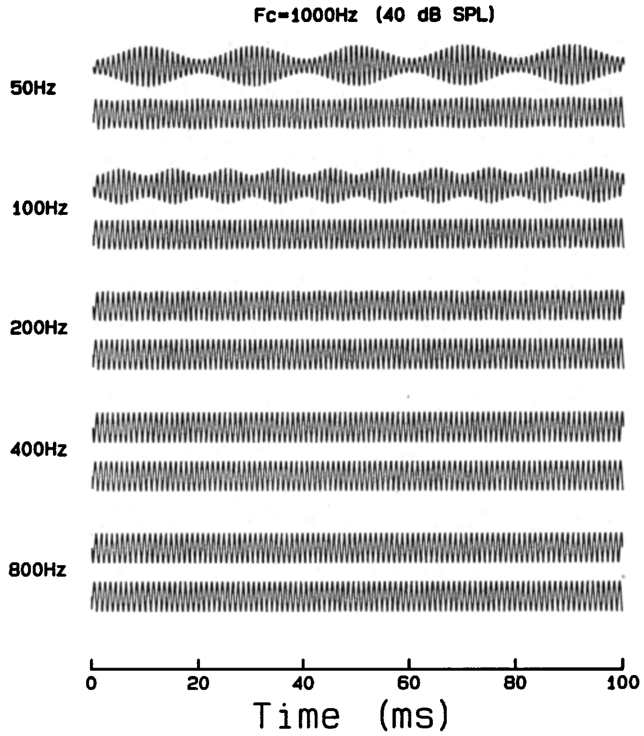

FIG. 7. Pairs of SAM and QFM waveforms at 40 dB SPL, which have been processed by an auditory analyzing filter using filter slopes inferred from forward-masked psychophysical tuning curves (Nelson, 1991). At each modulation frequency, the top waveform is the SAM stimulus and the bottom waveform is the QFM stimulus.

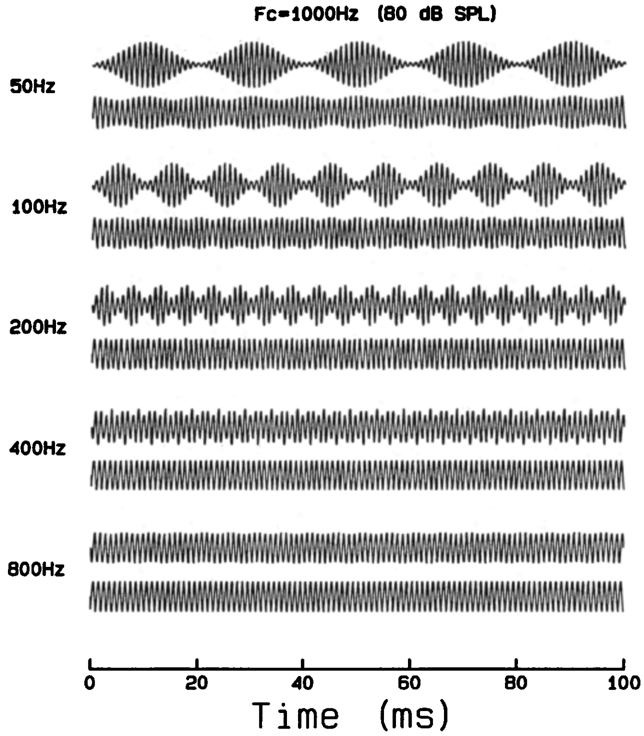

FIG. 8. Pairs of SAM and QFM waveforms at 80 dB SPL, which have been processed by an auditory analyzing filter using filter slopes inferred from forward-masked psychophysical tuning curves (Nelson, 1991).

$$S_{\mathrm{h}}=0.3852e^{(-0.0068L_{\mathrm{tip}})}. \tag{5c}$$

With the parameter estimates provided by Eqs. (5) we can calculate the excitation waveforms that would result from SAM and QFM stimuli passed through an auditory filter with slopes (in dB per Hz) inferred from psychophysical tuning curves.

At low stimulus levels, say 40 dB SPL, the low-frequency slope is steeper than the tail slope. Because of this the low-frequency slope has a greater effect on SAM and QFM waveforms than the tail slope. Consequently, S_1 is set to S_{lf}, with $S_2=S_{\mathrm{hf}}$, and the relative levels of the three components of 100% SAM and QFM are determined from Eq. (4). Figure 7 shows the waveforms that would result at 40 dB SPL with this auditory filter. Notice that at a modulation frequency of 50 and 100 Hz the amplitude envelope of the SAM waveform is no longer modulated by 100%, but the modulation is still sufficient so that the SAM waveform is easily distinguished from the QFM waveform by its envelope. For modulation frequencies at 200 Hz and above, the SAM and QFM waveforms are indistinguishable. Attenuation of the side bands by the auditory filter has removed any modulation so that both waveforms are essentially sinusoidal at the carrier frequency. This is qualitatively consistent with the mean $\mathrm{CB}_{\mathrm{phs}}$ of 118 Hz obtained at 40 dB SPL in the present study (Table I).

At high stimulus levels, say 80 dB SPL, the tail slope is steeper than the low-frequency slope. Therefore, S_1 was set to S_{tl} for our calculations. Figure 8 shows the waveforms that result at 80 dB SPL. In this case the SAM and QFM waveforms are easily distinguished by their envelopes for modulation frequencies from 50 to 400 Hz. Only at 800 Hz are the waveforms indistinguishable. This is also consistent with the mean $\mathrm{CB}_{\mathrm{phs}}$ of 414 Hz obtained at 80 dB SPL in the present study.

Goldstein's (1967a) psychophysical model assumed that SAM and QFM waveforms are indistinguishable when the power ratio (P) of the excitation envelope maxima is smaller than some threshold value. This is expressed in Eq. (6) [derived from Goldstein's Eq. (11)], which describes critical bandwidth for phase discrimination, given a threshold power ratio (P) and the low-frequency slope of the auditory filter (S_1):

$$\mathrm{CB}_{\mathrm{phs}}=\frac{\cosh^{-1}[1/(P-1)]}{0.115S_1}, \tag{6}$$

where

$S_1=S_{\mathrm{lf}}$, for $L_s<74$ dB SPL,

$S_1=S_{\mathrm{tl}}$, for $L_s>74$ dB SPL.

With the level dependencies of the low frequency and tail slopes of the auditory filter from Eqs. (5), $\mathrm{CB}_{\mathrm{phs}}$ as a function of overall level of the complex in dB SPL (L_s) is predicted from Eq. (6). Predictions were based on $S_1=S_{\mathrm{lf}}$ for levels between 30 to 74 dB SPL, and $S_1=S_{\mathrm{tl}}$ for levels above 74 dB SPL. The power ratio at threshold was 1.5 dB ($P=1.41$).

The model predictions from Eq. (6) are shown by the dashed line in Fig. 5. They indicate the $\mathrm{CB}_{\mathrm{phs}}$ values expected for envelope detection following filtering by a level-dependent auditory filter. The predictions are in good agreement with the mean data from the present study (Table I). The CB_{phs} values predicted from the low-frequency slope of the forward-masked tuning curve increase with stimulus level up to about 74 dB SPL. At 74 dB SPL the low frequency and tail slopes taken from fits to Nelson's (1991) data are equal. Above 74 dB SPL the tail slope is actually steeper than the low-frequency slope. Therefore, the tail slope determines CB_{phs} predictions, which follow the general trend of the mean data at 80 and 100 dB SPL.

With an experiment that involved the discrimination of phasic and antiphasic SAM envelopes at 50 dB SL in amplitude modulated noise, Buunen (1975) demonstrated that envelope detection alone is most likely responsible for phase discrimination of waveforms with modulation frequencies up to about 120 Hz at 1000 Hz. That agrees fairly well with the limits of phase discrimination at 50 dB in Fig. 5 predicted by the filter model.

Although the predictions of the filter model are representative of the geometric mean CB_{phs} values given in Table I, they do not represent, as well, the CB_{phs} values exhibited by the best performing subjects at all levels. Above 70 dB SPL, the individual values all cluster together and are reasonably well represented by the simple filter-model predictions. Below 70 dB SPL, the range of individual values is much larger. At 40 and 60 dB SPL, nine of the eighteen CB_{phs} values are less than those predicted by the filter model, which could easily be due to steeper filter slopes than those used here for prediction, so the filter model is adequate for these values. However, the remaining CB_{phs} values, particularly the five largest values, are greater than might be predicted by the filter model without any additional assumptions. In some cases they are more than an octave wider.

If the filter model parameters truly indicate the limits of performance with a simple envelope detector and a level-dependent auditory filter, the existence of CB_{phs} values much wider than predicted suggest that other contributing factors or auditory processes must be operating. One possibility is that the power ratio at threshold varies among individuals and perhaps even with level. Another is that filter slopes vary among individuals. These factors could account for some of the smaller differences among individuals seen in the present study, perhaps those seen above 70 dB SPL, but it is unlikely that they account for CB_{phs} values that exceed the prediction by an octave. Another possibility is that the level-dependent auditory filter may also introduce some relative phase shifts between components. This might account for narrower bandwidths than predicted, where phase ambiquities could lead to poorer performance at selective modulation frequencies, as seen at 40 dB SPL for subject KS(L), but the phase ambiquities should not lead to superior performance than predicted by the filter model.

From this we conclude that envelope detection following level-dependent auditory filtering can adequately predict the general increase in CB_{phs} with level, including asymptotic performance at high levels. This is primarily due to changes in the low-frequency side of the auditory filter with level. However, below 70 dB SPL, performance from the best subjects was much better than predicted. These departures from prediction suggest that the filter model may be an oversimplification, and that other mechanisms are involved as well. One such mechanism is the production of combination tones in the auditory periphery, which modify the internal excitatory waveshape (or internal spectrum).

B. Audibility limits for combination tones and SAM/QFM discrimination

Goldstein (1967b) and Smoorenburg (1972) have shown that combination tones of the first order can be perceived as long as the frequency separation between two tones is not too large. When the frequency separation between tones exceeds some limit, combination tones are no longer perceived. Smoorenburg also showed that the audibility limits for perceiving combination tones depended upon the overall level of the primaries. His subjects listened to equal-amplitude two-tone complexes as the frequency separation between components was reduced. That frequency separation where subjects just began to hear the combination tone rising in pitch was taken as the lower limit of combination-tone perception. Through pitch matching it was determined that their subjects heard the first order combination tone, or cubic difference tone (CDT). The audibility limits for CDT perception at 1000 Hz (the geometric mean across three subjects from his Fig. 1) are shown in Fig. 5 by the unfilled stars. Audibility limits for CDT perception measured in the same way by Buunen (1975) agree. Data from one of his subjects are shown by the unfilled square in Fig. 5.

Thus for modulation frequencies less than those indicated by the unfilled stars in Fig. 5, CDTs are likely to be heard by normal-hearing listeners. All of the CB_{phs} values obtained in the present study are well within the audibility region for CDT perception. So it could be that cubic difference tones influenced all of the CB_{phs} measures shown here, although, more likely, they influenced the data from only our best performers at lower levels. For example, at 40 and 60 dB SPL, the CB_{phs} values exhibited by our best subjects tended to be slightly smaller than, but parallel to, the CDT limits for two tones. This strongly suggests that the phase discrimination thresholds from the best subjects, those with CB_{phs} values well above the dashed line at 40 and 60 dB in Fig. 5, may well have been influenced by CDTs, and not simply by attenuation of the sidebands by the auditory filter.

This coincidence between the audibility limits for perceiving CDTs with two-tone complexes and the audibility limits for discriminating phase effects with three-tone complexes is necessary for a causative relation to exist between CDTs and phase effects, but is not sufficient. It does not define the mechanism or mechanisms by which CDTs influence phase perception. Further discussion might clarify some of those mechanisms.

C. Possible mechanisms

1. Phase-dependent internal spectrum

The CDTs of primary interest are of the form $2F_c - F_u = F_l$. They are produced by interactions between the center-frequency and upper-frequency component of the SAM/QFM complex, and have a frequency equal to that of the lower-frequency component of the complex. Buunen has described how the existence of these CDTs might influence the discrimination between SAM and QFM waveforms (Buunen and Bilsen, 1974; Buunen *et al.*, 1974, Buunen, 1975). The internal CDT at F_l has a magnitude and phase, which can interact with the external acoustic component at F_l to produce an *internal spectrum* that is different from the spectrum introduced into the ear. For modulation frequencies between about 120 Hz (at 1000 Hz) and the critical modulation frequency, Buunen proposed that phase discrimination is based upon differences in the phase-dependent internal spectrum. In support of that hypothesis, Buunen and Bilsen (1974) reported data from three subjects who showed a similar audibility range for phase effects and CDTs across test frequencies from 500 to 4000 Hz. Data from one of those subjects (JR) at 1000 Hz is shown in Fig. 5 by the open square for CDT audibility limits and by the filled square for the limits of phase perception. They supported their internal spectrum hypothesis further with forward-masking data that showed phase-dependent changes in forward masking at F_l but not at F_c or F_u. Results from three subjects showed that the amount of forward masking at F_l can vary by up to 10 dB as the phase of F_c is varied over 90° and the phase of F_u is held constant. Under the same conditions, forward masking at F_c and F_u did not change. Assuming the slope of the growth of forward masking for their conditions is around 0.5, their data suggest that the internal spectral component at F_l varied over a range as large as 20 dB for a 90° phase change in F_c.

These findings provide convincing evidence that a phase-dependent internal spectrum is produced by interactions between an internal CDT component at $2F_c - F_u$ and the external component at F_l, but exactly how those interactions manifest themselves in a perceptible cue to the nervous system is not obvious. Several possibilities present themselves.

2. Intensity discrimination at F_l

One possible cue for discriminating between SAM and QFM stimuli might be an increment or decrement in the loudness of the internal spectral component at F_l. Consideration of the results of previous CDT cancellation experiments provides some insight into the likelihood that intensity discrimination of the internal spectral component at F_l is a realistic cue.

Assuming that the power increment required for intensity discrimination is 1.0 dB or greater, the internal CDT level necessary to produce that power increment would have to be no less than 20 dB below the level of the acoustic component at F_l. For two-tone complexes, CDT level decreases as frequency separation between components increases. Data from Goldstein (1967b) indicate that the slope is about -0.125 dB/Hz. Allowing for possible overestimations of CDT levels with the cancellation technique by subtracting 6 dB from Goldstein's data in his Fig. 17, we see that this condition might exist for frequency separations less than about 200 Hz. For frequency separations greater than this, CDT level would be too low to produce a detectable power increment. Examining Table I, we see that at 40 dB SPL all but two of the CB_{phs} values are consistent with this power increment concept ($CB_{phs} < 200$ Hz). At 60 dB SPL about half of the values are less than 200 Hz. At 80 and 100 dB SPL, all of the CB_{phs} values are larger than 200 Hz. From this it does not seem likely, especially at the higher stimulus levels, that the significant cue for discriminating QFM from SAM waveforms is a power increment of the internal spectral component at F_l.

Furthermore, if the loudness of one component were to be the principal perceptual cue governing SAM/QFM discriminations, such loudness cues should only be available when the individual components are resolved into separate critical bands. Generally individual components are not resolved until they are separated by at least 20%, and then, according to the reasoning above, the CDT magnitude would be too low to produce detectable power increments. This frequency-resolution problem offers an argument against intensity discrimination at F_l as a usable cue.

3. CDT-induced envelope cues

The changes in the internal spectral component at F_l, caused by interactions between the internally generated CDT and the acoustic component at F_l, should also produce significant changes to the internal excitation waveform. The envelope of that internal excitation waveform could then serve as the effective cue to an envelope detector. In this case, the appropriate cue to discriminating SAM and QFM waveforms would still involve some type of envelope detection, such as the envelope power ratio described earlier, even though CDTs are involved in producing that cue. Such a CDT-induced envelope cue would avoid the frequency resolution problem encountered with the intensity-discrimination cue described above, since all three components must excite the same auditory filter to generate the CDT-induced envelope cue. Increases in the audibility limits with stimulus level are then explained by decreases in the slopes of the low-frequency side of the auditory filter with stimulus level. A measure of critical modulation frequency for phase discrimination would still reflect the limits of tonal interaction that are provided by the auditory filter. This is because the magnitude of the CDT is also subject to auditory filtering (Goldstein. 1967b; Smoorenburg, 1972).

This CDT-induced envelope cue also offers an explanation for the individual differences among subjects seen at 40 and 60 dB SPL in Fig. 5. Notice in Fig. 3 that at 40 dB SPL, subject KS demonstrated a nonmonotonic phase discrimination function. The subject lost the discrimination cue between SAM and QFM at a modulation frequency around 62 Hz, where performance dropped to chance. Then, as modulation frequency increased further, performance rose to 100% again until it dropped to chance for modulation frequencies above 128 Hz. One explanation might attribute the drop in performance around 62 Hz to the internal spectrum resulting from interactions between the CDT component and the external acoustic component at F_l. It is possible that the CDT phase was such that the 90° phase shift on F_c, which is equivalent to a 180° phase shift on F_l, produced an internal spectrum that differed little at 62 Hz because the phase of the CDT was near 90°. In which case, the vector sum between the internal CDT and the external component would yield phases for the internal spectral component at F_l of 45° and 135°, respectively, for F_c phases of 0° and 90°. The phase shift of 90° would be ambiguous because the envelopes for F_l phases of 45° and 135° are identical. As modulation frequency increased, the phase of the CDT component increased, and the 180° phase shift on the external component produced a vector sum at F_l that changed the internal excitation envelope sufficiently to be discriminated by the auditory system. Finally, above a modulation frequency of 128 Hz, the auditory filter removed all interactions between components and performance fell to chance levels. Such a phase shift in the CDT with increased modulation frequency is supported by cancellation data from Goldstein (1967b), where CDT phase increased fourfold for a doubling of modulation frequency (see his Fig. 13). At low stimulus levels this type of outcome is more likely than at high stimulus levels because the sharper filtering at low levels may result in a larger CDT phase shift with modulation frequency.

This explanation for individual differences among subjects is also supported by the existence of large differences among subjects in the apparent phase of the internal spectrum (Buunen and Bilsen, 1974; Buunen *et al.*, 1974). For three-component complexes similar to those used here, they showed that the apparent phase of the internal F_l component exhibited a large variation across subjects. Minimum levels of the internal F_l components were about 30°, 120°, and 150° in three different subjects. This variation across subjects was not only seen in cancellation experiments (Buunen *et al.*, 1974), but also in forward-masking experiments that preclude the criticisms common to cancellation experiments (Buunen *et al.*, 1977). If the auditory system were linear, and the internal excitation envelope were only determined by the output from the auditory filter, one would not expect such large individual differences in the phase of the internal component across subjects, nor would one expect to find discontinuities in the phase discrimination function of only some subjects, as seen in Fig. 3. It is likely that those discontinuities were due to large differences in the phase of the F_l component in the internal spectrum.

Similar nonmonotonic phase discrimination functions were reported by Nelson (1978). He presented 90° phase shifts in the center component of a three-tone complex to his subjects. In contrast to the phase discrimination functions shown here, which have a reference phase of 0°, Nelson held modulation frequency constant and varied the reference phase of the comparison complex from 0° to 180°. He reported that not all of the subjects performed best with the 0° vs 90° phase comparison. Some of the subjects did better for other reference phases, for example, 30° vs 120°. Although his results were only preliminary, they are consistent with the existence of an internal spectrum whose phase varies across individual subjects.

From this discussion, it appears very likely that combination tones influenced the measures of CB_{phs} presented here. It is clear, over the entire range of levels investigated, that combination tones are present at modulation frequencies corresponding to CB_{phs}, but the exact nature of how those combination tones influence measures of CB_{phs} is complex and not completely resolvable with the present data. The CDT-induced envelope cue is an attractive concept for understanding how SAM and QFM waveforms are distinguished. However, for any exact predictions to be accomplished, we need to know the magnitude and phase of the cubic difference tone at all levels and modulation frequencies, in each individual ear tested. These data are not available without extensive further research.

Regardless of the exact mechanism by which combinations tones affect the perception of phase changes, it appears that the audibility range of combination tones and the limits of CB_{phs} are limited by the same underlying process: nonlinear (level-dependent) auditory filtering.

IV. CONCLUSIONS

The present research confirms, and extends to higher stimulus levels, previous work showing increases in the critical modulation frequency for phase discrimination as a function of the level at which complex stimuli are presented to the ear. An objective measure of critical bandwidth for phase discrimination (CB_{phs}) was introduced. CB_{phs} increased with stimulus level between 40 and 80 dB SPL and remained relatively constant above 80 dB SPL. The increase in CB_{phs} with level between 40 and 80 dB SPL is well described by Eq. (2) with $k=34$ and $m=0.136$. Model predictions using auditory filter slopes derived from forward-masked tuning curves can adequately account for the general increase with level of the geometric mean CB_{phs}, up to 80 dB SPL. They can also account for the asymptotic behavior of CB_{phs} above 80 dB SPL. Comparisons between CB_{phs} and the audibility region for combination tones, from previous studies, indicate that combination tones are strongly implicated in CB_{phs} measures. It is suggested that level effects and individual differences can be understood by assuming that SAM and QFM discrimination is based upon the internal excitation envelope produced by the interaction between external acoustic components and internally generated cubic difference tones, both limited by a level-dependent auditory filter. CB_{phs} is interpreted as an objective measure of the largest frequency separation between spectral components for which tonal interaction can be demonstrated within the cochlea.

ACKNOWLEDGMENTS

This research was supported by grants DC-00110 and DC-00149 from NIDCD. The author expresses his appreciation to Nancy Markman and Anna Schroder for their assistance during data analysis.

[1]These four tuning curves were obtained using four different probe levels but, as shown by Nelson and Freyman (1984), similar changes in tuning curve shapes can be obtained by varying the delay time between masker and probe tone, which indicates that the levels of the maskers, not the levels of the probe tone, determine tuning-curve shapes.

Buunen, T. J. F. (**1975**). "Two hypotheses on monaural phase effects," Acustica **34**, 98–105.

Buunen, T. J. F., and Bilsen, F. A. (**1974**). "Subjective phase effects and combination tones," in *Facts and Models in Hearing*, edited by E. Zwicker and E. Terhardt (Springer-Verlag, New York), pp. 344–352.

Buunen, T. J. F., Festen, J. M., Bilsen, F. A., and van den Brink, G. (**1974**). "Phase effects in a three-component signal," J. Acoust. Soc. Am. **55**, 297–303.

Buunen, T. J. F., ten Kate, J. H., Raatgever, J., and Bilsen, F. A. (**1977**). "Combined psychophysical and electrophysiological study on the role of combination tones in the perception of phase changes," J. Acoust. Soc. Am. **61**, 508–519.

Goldstein, J. L. (**1967a**). "Auditory spectral filtering and monaural phase perception," J. Acoust. Soc. Am. **41**, 458–479.

Goldstein, J. L. (**1967b**). "Auditory nonlinearity," J. Acoust. Soc. Am. **41**, 676–689.

Mathes, R. C., and Miller, R. L. (**1947**). "Phase effects in monaural perception," J. Acoust. Soc. Am. **19**, 780–797.

Nelson, D. A. (**1977**). "Phase discrimination and level effects," J. Acoust. Soc. Am. Suppl. 1 **61**, S50.

Nelson, D. A. (**1978**). "Monaural phase effects in listeners with sensorineural hearing loss: preliminary findings," ARO Abstracts, p. 13.

Nelson, D. A. (**1991**). "High-level psychophysical tuning curves: forward masking in normal-hearing and hearing-impaired listeners," J. Speech Hear. Res. **34**, 1233–1249.

Nelson, D. A., Chargo, S. J., Kopun, J. G., and Freyman, R. L. (**1990**). "Effects of stimulus level on forward-masked psychophysical tuning curves in quiet and in noise," J. Acoust. Soc. Am. **88**, 2143–2151.

Nelson, D. A., and Freyman, R. L. (**1984**). "Broadened forward-masked tuning curves from intense masking tones: delay-time and probe-level manipulations," J. Acoust. Soc. Am. **75**, 1570–1577.

Rosen, S. (**1984**). "Hyperacute monaural phase sensitivity in the hearing impaired," Br. J. Audiol. **18**, 257–258 (A).

Rosen, S. (**1986**). "Monaural phase sensitivity: frequency selectivity and temporal processes," *Auditory Frequency Selectivity*, edited by B. C. J. Moore and R. D. Patterson (Plenum, London), pp. 419–427.

Rosen, S. (**1987**). "Phase and the hearing impaired," in *The Psychophysics of Speech Perception*, edited by M. E. H. Schouten (Martinus Nijhoff, Dordrecht).

Rosen, S., and Fourcin, A. (**1986**). "Frequency selectivity and the perception of speech," in *Frequency Selectivity in Hearing*, edited by B. C. J. Moore (Academic,London), pp. 373–487.

Rosen, S., and Smith, D. A. J. (**1988**). "Temporally-based auditory sensations in the profoundly hearing-impaired listener," in *Basic Issues in Hearing: Proceedings of the 8th International Symposium on Hearing*, edited by H. Duifhuis, J. W. Horst, and H. P. Wit (Academic, London), pp. 431–439.

Schroder, A. C., and Leek, M. R. (**1989**). "Frequency selectivity and monaural phase discrimination in hearing-impaired listeners," Assoc. Res. Otolaryngology Abstr. (12th Annual Mtg.), 303.

Smoorenburg, G. F. (**1972**). "Audibility region of combination tones," J. Acoust. Soc. Am. **52**, 603–614.