The role of distortion products in masking by single bands of noise

Please check the document version of this publication:

Link to publication

The role of distortion products in masking by single bands of noise

Marcel van der Heijden and Armin Kohlrausch
Institute for Perception Research (IPO), P.O. Box 513, 5600 MB Eindhoven, The Netherlands

Masking experiments with frozen-noise maskers were conducted to investigate the role of distortion products generated by the interaction between the components of a bandpass noise masker. In the first experiment, thresholds of a 900-Hz sinusoidal signal with a duration of 50 ms (10-ms ramps included) were measured in the presence of bandpass noise maskers ranging from 1 to 2 kHz. In all measurements the same 500-ms noise sample was used (frozen noise), presented at overall sound-pressure levels of 35, 50, 65, or 80 dB. The signal was temporally centered in the masker. Four subjects participated in the experiment. Threshold variations of more than 10 dB were observed on varying the signal phase. The pattern of threshold versus signal phase resembled a sinusoid; the signal phases at the minimum and maximum thresholds differed by about 180 deg. The phase pattern shifted with increasing masker level. The individual shift for the masker range of 45 dB was between 100 and 200 deg. The direction of this shift agreed with data on the phase of cubic difference tones as a function of the level of sinusoidal primaries. In a second experiment a large number of different frozen-noise samples were used as maskers in order to evaluate the generalizability of the phase effect. Two types of noise maskers were used: a low-frequency masker (0–800 Hz) and a high-frequency masker (1000–2000 Hz). For each noise sample, thresholds of a 900-Hz signal were measured for two signal phases, 0 and 180 deg. An analysis of variance showed that signal phase played a significant role for the high-frequency, but not for the low-frequency, masker. © *1995 Acoustical Society of America.*

PACS numbers: 43.66.Dc, 43.66.Mk, 43.66.Ba

INTRODUCTION

Pure tones and bands of noise show considerably different masking behavior. Many factors are likely to play a role in this difference. For instance, the use of beats as a detection cue can improve detectability of a sinusoidal signal when the masker is a pure tone, provided that the two frequencies are not too far apart. The inherent fluctuations of a noise masker hamper this cue (Egan and Hake, 1950).

The high-frequency slope of pure-tone masking patterns can be considerably affected by the generation of distortion products *below* the masker's frequency. The most important role is usually played by odd-order distortion products of the type $f_l - n(f_h - f_l)$, where f_l and f_h are the frequencies of the lower and higher primaries, respectively, and n is a small positive integer (Goldstein, 1967). In a limited range of frequencies above the masker frequency, masked thresholds appear to be determined by detection of these distortion products rather than by a direct detection of the signal. This artifact is revealed by irregularities in the upper slope of the tone-on-tone masking pattern (Greenwood, 1961), which can usually be removed with the help of a low-pass noise of moderate intensity. The influence of distortion products is not restricted to pure-tone maskers, but is also found in the case of noise bands. The detection of the distortion products, however, depends on the masker bandwidth (Greenwood, 1971). This is a second cause of the different amounts of masking produced by noise bands and pure tones.

When the signal frequency is well above the masker frequency, a noise masker is usually less effective than a pure tone (Buus, 1985). Masked thresholds are affected by both bandwidth and statistics of the masker (Mott and Feth, 1986; van der Heijden and Kohlrausch, 1995). The "release from masking" observed in the case of a noise masker is usually attributed to the ear's ability to make use of minima in the temporal envelope of the masker (Buus, 1985).

Interestingly, the reverse effect occurs when the signal frequency is below that of the masker. In this situation a band of noise of sufficient bandwidth generally produces *more* masking than a pure-tone masker of the same intensity (Zwicker and Bubel, 1977; Mott and Feth, 1986; Glasberg and Moore, 1994). Both Zwicker and Bubel (1977) and Glasberg and Moore (1994) suggested that the masking by noise bands toward low frequencies is influenced by the generation of distortion products by components within the masker. Unlike in the situation described above, where distortion products facilitate the detection of the signal, in this scenario the distortion products are supposed to produce masking.

A. Masking produced by distortion products

Greenwood (1971) demonstrated that combination products can produce masking. For maskers consisting of pairs of narrow-band stimuli he found local maxima in the masking pattern at frequencies below the masker spectrum. He argued that "...presenting two bands of noise as primary stimuli, which then generate combination bands, is in principle no different than the presentation of only one band [which] should generate an overlapping series of combination components" (Greenwood, 1971, p. 525). For this overlapping series Greenwood coined the term *combinational aggregate.*

There are, however, some problems associated with the implicit mathematical reasoning behind the "combinational aggregate."

In Greenwood's procedure, the distortion of an arbitrary collection of primaries is constructed by a simple *summation* of distortion products that would be produced by single pairs of components when presented in isolation. The validity of this construction for an arbitrary nonlinear process is not obvious; in fact, such splitting-and-adding schemes can only be safely applied to linear systems, in which "the effect of the sum" is by definition equal to "the sum of the effects." In both Greenwood (1971) and Zwicker and Bubel (1977) this assumption underlies their attempts at estimating the level of the combinational aggregate and its masking potency. Greenwood sought support for it by inspecting data obtained with various combinations of *three* narrow-band maskers, from which he concluded "that the generation of combination components by any pair of primaries is not precluded by the presence of any of the other primaries" (Greenwood, 1971, p. 525). But "not precluded" is not the same as "unaffected," and the generalization from three isolated masker components to an infinite number of adjacent ones is not self-evident.

There is experimental evidence that the simple summation of combination bands produced by component pairs is an inadequate procedure; Lutfi (1983) observed that the threshold of a signal at the frequency of the cubic combination band ($2 f_l - f_h$) produced by two narrow-band primaries actually decreased when the spectral gap between the two primaries was filled with additional noise.[1] The construction of the aggregate proposed by Greenwood (1971, p. 532ff.) is also incompatible with the simple compressing type of nonlinearity that has been proposed as the origin of odd-order difference tones in the auditory system (Smoorenburg, 1972). For an analysis of the spectral transformations resulting from nonlinear processing of more complex stimuli, the reader is referred to de Boer (1976).

Although the data of both Greenwood (1971) and Zwicker and Bubel (1977) agree with the hypothesis of masking by combination products, they do not present conclusive evidence for the aggregate. In the case of distinct primary bands, masking by combination bands is evidenced by the occurrence of irregularities in the masking patterns. Such evidence is absent in masking patterns for single noise bands. The low-frequency portion of the combinational aggregate produced by a single masker band (i.e., those distortion components that are below the lower cutoff frequency of the primary band) can safely be expected to show a high-pass character; the intensity of combination products decreases with increasing order. An aggregate with such a spectral shape would cause an "extra" downward spread of masking that could equally well be due to the limited spectral resolution of a perfectly linear mode of processing. For this reason masking patterns are not a very adequate tool for evaluating the role of distortion products when the masker consists of a single band of noise. In the present study we present and evaluate a new method for examining the role of distortion products in masking produced by a single band of noise.

B. Phase-dependent signal–masker interactions

The approach of the present study is based on the fact that signal–masker interactions can affect the audibility of a tonal signal in the presence of a broadband noise (Pfafflin and Mathews, 1966). In order to be able to observe this effect in a masking experiment, the same sample of masker noise (reproducible or "frozen" noise) has to be presented in each observation interval of the measurement, and the temporal placement of the signal within the masker has to be identical in all test intervals. On changing the (monaural) signal phase, thresholds have been observed to vary in a sinusoidal fashion, the threshold differences being larger than 10 dB in some cases (Hanna and Robinson, 1985; Langhans and Kohlrausch, 1992; von Klitzing and Kohlrausch, 1994). This effect is not surprising since, as Pfafflin and Mathews (1966, p. 344) put it, "by changing the phase of the signal relative to the samples making up the noises, the energy increment produced by the signal is altered."

A crucial condition for phase effects to occur is of course that the signal and portions of the masker have a meaningful relative phase, i.e., that their spectra overlap. More precisely, in order for two stimuli to show a phase-dependent interaction, the smallest spectral gap between them should not exceed the reciprocal value of the duration of the overlap. In the case of a larger spectral distance between the signal and the masker, one would not expect a systematic effect of signal phase variation on the masked thresholds.[2] If a phase effect should nevertheless be observed in this situation, it would point to the existence and interaction of "secondary" stimuli which overlap spectrally. For instance, the combinational aggregate produced by a single band of noise could be responsible for (part of) the masking of a lower frequency signal. The reverse case is also conceivable: If signal–masker interactions cause distortion products that actually determine the detectability of the signal, then these distortion products could interact with spectrally overlapping portions of the masker. In either case, phase effects would reveal an essential role of distortion products in the masking process. Once such phase effects have been observed, an inspection and manipulation of the stimulus parameters will contribute to an understanding of the generation of distortion products and their importance in different masking situations. The present study intends to explore the possibilities of this approach.

I. EXPERIMENT 1: PHASE EFFECT FOR A SINGLE FROZEN-NOISE SAMPLE

A. Method

1. Stimuli

A single 500-ms sample of Gaussian noise was used as a masker throughout this experiment. It was bandlimited between 1 and 2 kHz and provided with 20-ms Hanning ramps. Masked thresholds were measured at masker levels of 35, 50, 65, and 80 dB SPL, corresponding to spectrum levels of 5, 20, 35, and 50 dB/Hz, respectively. Bandlimited Gaussian noise was produced as follows. First, a 500-ms buffer of wideband Gaussian noise was obtained by drawing independent samples from a Gaussian distribution. A discrete Fourier

transform was applied to this buffer, which led to a spectrum with a spacing of 2 Hz between the components. The spectral components below 1 kHz and above 2 kHz were set to zero and an inverse Fourier transform yielded a 500-ms buffer of bandlimited Gaussian noise. In this way, the long-term spectrum of the cyclic noise buffer had infinitely steep spectral edges. The steepness of the spectral edges of the actual noise maskers, as presented in each interval, was only limited by their duration of 500 ms and the use of Hanning ramps of 20 ms.

The signal was a 900-Hz sinusoid with a total duration of 50 ms, temporally centered in the masker. The signal was provided with 10-ms Hanning ramps, and its starting phase was varied in 45-deg steps. All stimuli were digitally generated at a 32-kHz sampling rate and played out using the built-in 16-bit D/A converters of a Silicon Graphics Iris computer.

The stimuli were presented diotically via a Telephonics TDH 49 headset mounted in fluid-filled circumaural cushions. The earphone had previously been calibrated by means of a probe microphone placed at the ear canal entrance of several subjects. The response was flat ± 3 dB over the range from 500 to 6000 Hz as measured at the ear canal entrance. Distortion was measured acoustically for a pair of primaries at 1000 and 1200 Hz, presented at a level of 85 dB SPL per tone. The level of the intermodulation products was at least 70 dB below the primary level.

2. Procedure and subjects

Masked thresholds were determined using a three-interval forced-choice adaptive procedure (Levitt, 1971). Each trial consisted of three 500-ms observation intervals separated by 200 ms of quiet. The masker occurred in all three intervals. The signal occurred randomly but with equal probability in one of the three intervals. After the subject's response had been collected, a 300-ms pause preceded the next trial. Correct-answer feedback was provided on a computer screen. Each trial block began with the signal about 20 dB above masked threshold. After two consecutive correct responses at the same signal level, it was decreased and for each incorrect response it was increased. This procedure tracks the 70.7%-correct point of the psychometric function. The step size was 8 dB at the beginning of each block, was reduced to 4 dB after the second reversal, and to 2 dB after the fourth reversal. Using 2-dB steps, ten more reversals were obtained. The threshold for a block was estimated by taking the median of the signal levels of these ten last reversals. The thresholds reported in this paper are the averages of three single threshold estimates.

The subjects were tested in a single-walled sound-attenuating booth placed within a larger sound-attenuated room. Four normal hearing subjects, A–D, participated in the experiment. Subjects C and D were the authors. All subjects had extensive experience in listening tasks.

3. Pilot study: Temporal placement of the signal within the masker

In the case of a frozen broadband noise masker, the effect of signal starting phase on masked threshold is both well

established and well understood in terms of masker–signal interactions (Hanna and Robinson, 1985). The size of the effect, however, can vary considerably between different noise samples. This is not surprising, because some noise samples can show phase modulations that make the masker–signal interactions essentially incoherent, thereby preventing the occurrence of phase effects.

In the present study we are trying to find out whether phase effects occur under off-band conditions. It is therefore desirable to select a frozen noise sample so as to optimize the expected phase effect. From on-frequency masking data of Langhans and Kohlrausch (1992) it appears that the lowest thresholds produced by the "phase-sensitive" noise samples are generally lower than the thresholds produced by noise samples that yield no phase effect. That is, the phase effect can produce a *release from masking* compared to masking by "phase-insensitive" noise samples. This suggests that, among a set of noise samples, those samples that yield the lowest threshold for an arbitrary fixed phase value are likely to show an effect of signal starting phase. The utility of this idea to select noise samples was confirmed by our previous experience with masking by frozen noise in an on-frequency situation.[3]

We preceded the actual measurements by a pilot experiment for selecting a noise sample. Different noise samples were realized by successive 50-ms cyclic shifts of the masker: The noise stimulus was shifted (rotated) within a 500-ms cyclic buffer. In this way, the central 50-ms parts of all ten rotated versions of the masker (the parts that temporally overlapped with the signal) had nothing in common, since the shift was 50 ms. The ten versions of the masker may hence be considered practically independent noise samples.

Masked thresholds of a 900-Hz signal were measured as described above for ten different placements of the signal within the noise buffer. The masker level was 50 dB SPL. For each condition the threshold was measured twice. Data
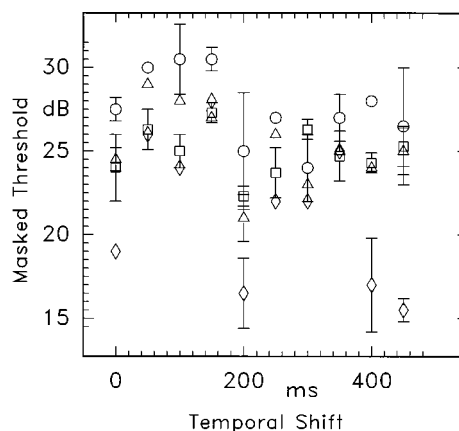
FIG. 1. Results of the pilot experiment, in which thresholds of a 50-ms tonal signal at 900 Hz were measured, masked by a 500-ms frozen sample of bandpassed noise ranging from 1 to 2 kHz. The error bars show $\pm$ one within-subject standard deviation. The cyclic masker noise buffer was shifted in 50-ms steps as indicated along the abcissa. Different symbols indicate data of different subjects: subjects A (circles), B (squares), C (triangles), and D (diamonds). Each symbol presents the average of two data points.
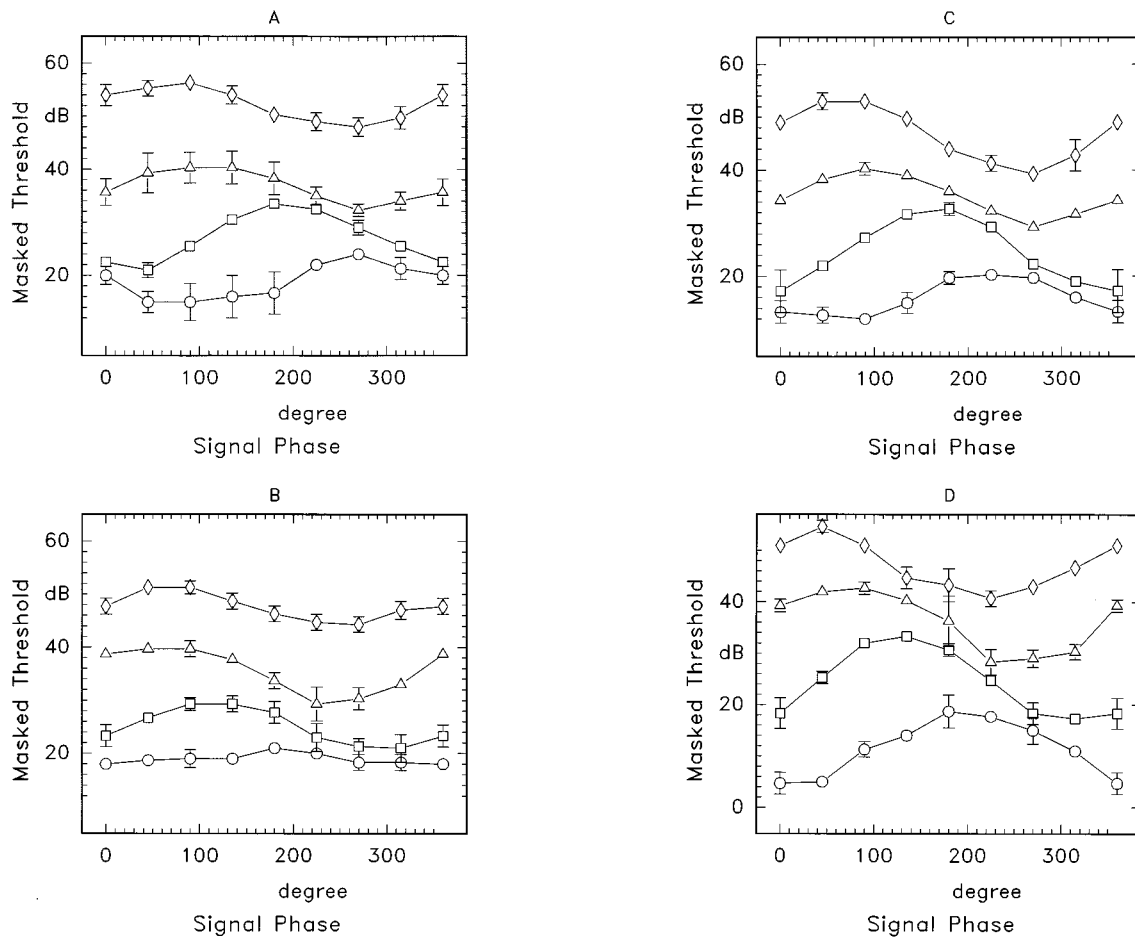
FIG. 2. Masked thresholds of a 50-ms tone at 900 Hz as a function of signal starting phase, which was varied in 45° steps. The masker was a 500-ms frozen-noise sample, bandpassed from 1 to 2 kHz. Panels A–D show the results for subjects A–D, respectively. In each panel, thresholds for four different masker levels have been plotted: 35 (circles), 50 (squares), 65 (triangles), and 80 dB SPL (diamonds). Thresholds measured with the same masker level have been connected by lines; for visual clarity the 0° data have been plotted twice (at 0° and at 360°). Each symbol presents the average of three data points; error bars indicate ± one standard deviation.

were collected from the same four subjects that participated in the main experiment.

The results are shown in Fig. 1. The thresholds varied over a range of 16 dB. Although there were considerable intersubject differences, there seemed to be a systematic effect of shift as well. This was confirmed by a two-way analysis of variance (shift versus subject; both treated as fixed effects), which revealed significant effects ($p<0.0001$) of both shift and subject. The interaction between shift and subject was found to be significant as well ($p<0.0004$).

Since the shift of 200 ms led to the lowest average thresholds, this shift was used in the main experiment.

B. Results of experiment 1

Figure 2 shows the thresholds as a function of signal starting phase for subjects A–D (panels A–D). In each panel thresholds for four different masker levels are shown; different symbols have been used for each masker level. Lines connect thresholds measured using the same masker level; for visual clarity the threshold for 0° signal phase has been plotted twice (at 0° and 360°). The data of all subjects show an effect of signal phase; the thresholds vary in a sinusoidal fashion with one maximum and one minimum, separated by

approximately 180°. The difference between the maximum and the minimum varies between 3 (subject B, 35-dB masker) and 16 dB (subject D, 50-dB masker). The phase effect can be observed for all masker levels (35–80 dB SPL). There is an effect of masker level on the phase effect: Apart from an expected overall increase of thresholds with increasing masker level, the patterns are also shifted to the left. A further analysis of the data is presented in Figs. 3–5.

Growth-of-masking functions (threshold versus masker level) for each of the four subjects were obtained by averaging the thresholds belonging to each pattern in Fig. 2 over all phase conditions. In this way, a curve relating average threshold to masker level was obtained for each subject. The curve in Fig. 3 is the average of the four individual growth-of-masking functions; the error bars indicate the intersubject standard deviation. The threshold increases by 31.8 dB for a 45-dB increase in masker level. This corresponds to a slope of 0.71 dB/dB.

As a measure of the size of the phase dependency we calculated the standard deviation of subsets of the thresholds that only differ in phase (the first eight points of each of the patterns of Fig. 2). This resulted in a curve relating "spread caused by phase" to masker level for each subject. The av-
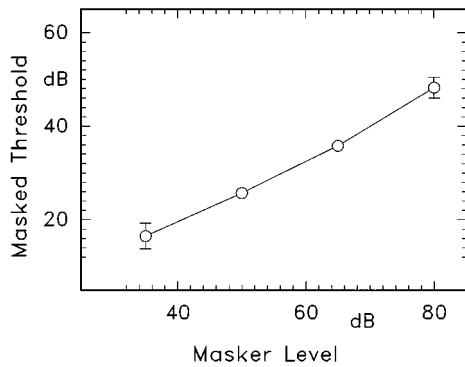
FIG. 3. Average growth-of-masking function (threshold versus masker level). The thresholds were obtained from the data of Fig. 2 by averaging the data for each masker level over all signal phase values and over all subjects. The error bars indicate ± one intersubject standard deviation.

erage of these four curves is shown in Fig. 4; the error bars indicate the intersubject standard deviation. The phase effect was found to be maximal for a 50-dB masker level. A two-way analysis of variance (masker level versus subject[4]), however, showed that the effect of masker level was not significant ($p > 0.05$).

In order to analyze the phase shift of the patterns of Fig. 2 with increasing masker level we calculated cross-correlation functions between the individual patterns of Fig. 2. In these calculations, the curves were treated as being cyclic (with a cycle of 360 deg). The analysis was performed in two steps: first a within-subject comparison, and next an across-subject comparison. For each subject the pattern for a 50-dB masker level was used as the reference. The shift of one threshold pattern relative to the reference pattern was defined as the phase value at which their cross-correlation function was maximal. In this way, the relative phases of the four patterns obtained using different masker levels were estimated for each subject. In order to be able to relate these phase values across the subjects, the same procedure was applied to the 50-dB data of all subjects, now using the 50-dB pattern of subject D as a reference.[5] The offsets that resulted from this procedure were added to the corresponding curves from the within-subject analysis.
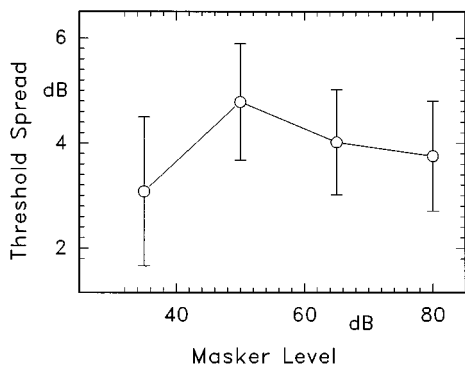

FIG. 4. The amount of phase-induced variation of thresholds, expressed by the standard deviation of the points contained in each pattern in Fig. 2, plotted as a function of masker level. Each data point presents the value obtained by averaging over all four subjects. The error bars indicate ± one intersubject standard deviation.
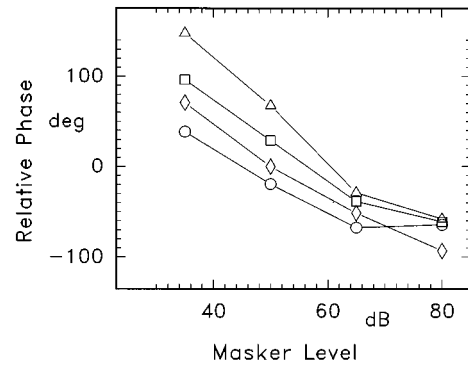

FIG. 5. Phase of the individual patterns of Fig. 2 relative to the 50-dB pattern of subject D, plotted as a function of masker level (see text). Different symbols are used for the four subjects: subjects A (circles), B (squares), C (triangles), and D (diamonds).

The resulting curves are plotted in Fig. 5. They show the phase of the individual patterns of Fig. 2 relative to the 50-dB pattern of subject D. It can be inferred from Fig. 5 that the phase of the patterns decreases monotonically with an increasing masker level. The total shifts over the 45-dB masker level range are 205, 106, 157, and 163 deg for subjects A–D, respectively. As the phase shift seems to saturate at the highest masker level for most subjects, we calculated the slopes of the curves for the lowest three masker levels only. These slopes were 3.9, 2.4, 3.0, and 2.7 deg/dB for subjects A–D, respectively.

C. Discussion

1. Masking by the combinational aggregate

Both the shape and the size of the patterns of Fig. 2 bear a close similarity to frozen-noise masking data measured under "on-frequency conditions," i.e., with a tonal signal placed within the spectrum of a broadband masker. Hanna and Robinson (1985) measured masked thresholds of a 100-ms tone at 500 Hz, temporally centered in a 150-ms reproducible noise with a flat spectrum extending from 100 to 3000 Hz. Ten different noise samples, presented at a spectrum level of 50 dB/Hz, were used as maskers. The signal starting phase was varied in 45-deg steps. Most of the threshold-versus-phase patterns reported by Hanna and Robinson (1985, their Fig. 3) have the same sinusoidal shape as the patterns of our Fig. 2. Threshold variation caused by signal phase ranged from a few to 12 dB. Langhans and Kohlrausch (1992) measured thresholds of 5-ms tones at 1 kHz, masked by broadband frozen noise. The masker had a spectrum level of 36 dB/Hz and a duration of 300 ms. They reported phase-induced threshold variations ranging from 6 to 13 dB. Again, the patterns presented in Figs. 1 and 2 of their paper bear a close resemblance to our observed patterns, shown in our Fig. 2.

In most studies reporting the effects of signal phase on masking by reproducible noise, the authors state that their data can be adequately explained by energy or envelope detection (Pfafflin and Mathews, 1966; Gilkey *et al.*, 1985; Hanna and Robinson, 1985; Langhans and Kohlrausch, 1992). This involves a straightforward evaluation of the en-

ergy or the envelope of that portion of the stimulus that is close to the signal frequency. Langhans and Kohlrausch (1992) argued that the effect can be well understood from direct interactions between the signal and neighboring masker components, provided one assumes that a temporary *decrease* of the envelope of the fluctuating masker is not detectable, whereas a sufficiently large temporary *increase* is detectable.[6]

In the present study, such a phase-dependent signal–masker interaction is excluded by the spectral separation between the two stimuli. The lowest masker components are still 100 Hz above the signal frequency, which means that their phase relative to the signal is "running around" three times during the 30-ms plateau of the signal burst. This incoherence between signal and masker means that the two have no phase relation, and that there cannot be any meaningful phase-dependent interaction between them. This implies that the phase effect found in experiment 1 can only be caused by the interaction of spectrally overlapping "secondary stimuli" resulting from nonlinearities. More precisely, the explanation of the present data demands a scheme in which either the signal or the masker (or both) are exposed to a nonlinear distortion prior to their merging into the input channel(s) of a detector. Any linear type of imperfect frequency resolution fails to account for the data, irrespective of the exact detecting mechanism.

The most probable explanation for the present stimuli is that the frozen-noise masker produces a "frozen combinational aggregate" of the type postulated by Greenwood (1971). An aggregate extending over the spectral region around the signal frequency would essentially create an on-frequency situation and hence explanations based on direct interaction would apply as before. The high degree of similarity between the present data and on-frequency data obtained in other studies supports the validity of this explanation.

On the basis of the on-frequency data of Hanna and Robinson (1985), we made a rough estimate of the spectrum level of the combinational aggregate around 900 Hz. Corrections had to be made for the differences in signal duration and frequency.[7] For our 65-dB masker, which had a spectrum level of 35 dB/Hz, the estimated spectrum level of the distortion products around 900 Hz is 19 dB/Hz. This estimate is not very precise, since the noise maskers used by Hanna and Robinson had a flat spectrum around the signal frequency, whereas the combinational aggregate can be expected to have a steep high-pass characteristic, which makes the direct comparison between their masking effects questionable.

We have attempted to compare this estimated spectrum level of the combinational aggregate with predictions of the constructions by Greenwood (1971, p. 532ff.) and Zwicker and Bubel (1977). As was described in the Introduction, these investigators proposed to estimate the aggregate by summing the distortion products caused by individual pairs of masker components. A closer look at this scheme reveals that the estimates obtained in this way are dependent on the *number* of masker components. This is easily seen by noting that the number of component pairs is quadratic in the number of components. If the overall level of the masker is kept

constant, a doubling of the number of masker components results in a halving of the power per component. Since the level of the odd-order difference tones is roughly linear in the overall level of the primary pair (Smoorenburg, 1972; Zwicker, 1979), the net result of doubling the number of components is a doubling of the power contained in the aggregate.[8] This curious property precludes the application of the scheme to a single noise band; here the decomposition in components is only an artificial construction, the details of which should have no impact on the derived result. If one were to identify the number of components with the number of Fourier components of the noise bands, then one would arrive at the unacceptable result that the level of the combinational aggregate grows in an unbounded manner with the duration of the noise stimulus, the number of Fourier components being linear in the duration.

In Zwicker and Bubel (1977), such problems were circumvented by only considering a group of five primaries after it had been experimentally observed that five irregularly spaced tones were sufficient in reproducing the low-frequency part of a masking pattern produced by a single noise band of critical bandwidth. Taking a magic number of five primaries per critical band, however, can hardly be called a satisfactory solution to the problems encountered in the estimation of the aggregate.

In the case of a pair of narrow-band noise primaries, the connection with two-tone data can be saved by regarding the two narrow bands as modulated sinusoids, and by assuming that the distorting process is fast enough to act at each instant as though the primaries were unmodulated, i.e., to produce a number of modulated difference tones. This "instant-wise" approach is basically different from the "component-wise" approach proposed by Greenwood (1971), and is not applicable to the wider noise bands used in the present experiment, in which the distinction between lower and higher primaries is lost.

2. Compatibility with known properties of combination products

The preceding analysis supports our view, also expressed in the Introduction, that even a detailed knowledge of combination tones (CTs) produced by pairs of primaries does not help much in predicting the exact form of distortion products of single noise bands. On the other hand, the fact that both types of distortion products are supposed to originate from the same nonlinear process suggests that their general properties should not be too different. In the following discussion, aspects of our data will be compared with corresponding aspects of odd-order CTs.

The 45-dB masker range for which the phase effects are observed agrees with the large dynamic range in which odd-order difference tones are observed. Goldstein (1967) and Smoorenburg (1972) reported the detection of cubic difference tones (CDTs) for primary levels as low as 20 dB SL. Zwicker (1979) reported CDTs for levels of 30–90 dB SPL per primary. In the same study, fifth-order difference tones were observed over a range of primary levels of at least 40 dB.

Zwicker's data on CTs produced by two primaries of equal level (this condition seems the most appropriate for a comparison with single-band distortions) show an almost linear increase of third- and fifth-order CTs with primary level in the range from 30 to 80 dB per primary. On the basis of such data one would expect the level of the combinational aggregate to increase linearly with the masker level, which contrasts with the shallow growth-of-masking function of the present experiment (0.71 dB/dB, see Fig. 3).

This expectation is based on a two-step argument: (1) The aggregate grows linearly with primary level, and (2) on-frequency masking shows a linear growth. Step 2 is based on a comparison with masking by spectrally flat broadband noise (Hawkins and Stevens, 1950). Possibly the expected steep high-pass character of the spectral portion of aggregate *below* the primary band causes a deviation from the perfect linear growth of a white-noise masker. It seems more likely, however, that step 1 is unjustified, as it contains a generalization from pairs of tones to bands of noise of exactly the type that is criticized in Sec. A of the Introduction. Lutfi (1983) found that the masked threshold of a signal at the CDT frequency of two primary bands actually *decreased* on filling the spectral gap between the masker primaries with an extra noise. Furthermore, Goldstein (1970) observed the effect of a third primary tone with frequency $2 f_2 - f_1$ on odd-order CTs produced by a pair of tones at f_1 and f_2; two different adjustments of the phase of this third tone (the level was kept constant) led to a complete cancellation of the CDT in one case and of the fifth-order CT in the second case. Such data show that the generation of distortion products is affected in a nonlinear way by the presence of components other than the primaries. The growth of the aggregate with the primary band level can therefore be very different from the linear growth observed for combination tones.[9] The shallow growth-of-masking function does not seem to be caused by the frozen-noise paradigm of the present study, as the slope of 0.71 of the present experiment agrees reasonably with the value of 0.77 reported by Glasberg and Moore (1994) for their most comparable condition, a (running) noise ranging from 1150 to 1550 Hz masking a 1-kHz signal.

The phase shift with increasing masker level, apparent from the raw data in Fig. 2, and in analyzed form in Fig. 5, has the same direction as and is comparable in size with the phase shift of CDTs with primary levels found in many studies. Zwicker (1979), using a cancellation procedure, found a downward phase shift in CDTs amounting to 90–120 deg for an increase in primary level of 60 dB, depending on the subject. For the fifth-order difference tone, a phase shift of about 200 deg over a 40-dB primary level range was found. Smoorenburg (1972), also using a cancellation procedure, found a downward phase shift of CDTs of about 360 deg for a primary level increase of 50 dB. Lutfi and Yost (1981), using a binaural paradigm to estimate CDT characteristics, found downward phase shifts of 50–150 deg for a 25-dB primary-level increase, depending on the subject.

A detailed comparison of the amounts of phase shift obtained in these different studies is meaningless, since the level-induced phase shift varies not only with the order of the CTs, but also with the frequency separation of the prima-

ries. There are, however, two qualitative similarities between our data and those obtained in many studies on CTs which are worth mentioning. First, in all three studies cited above, the downward phase shift saturates for high primary levels (Zwicker, 1979, Fig. 3; Smoorenburg, 1972, Fig. 3; Lutfi and Yost, 1981, Fig. 5). This saturation is also apparent in the data of three of the four subjects of the present study (see Fig. 5). Second, individual phase shifts, though conforming to the same trends, show a large variation across subjects (up to 100 deg over a total range of only a few hundred degrees). At the same time, the levels of the distortion product show little intersubject variation (only a few dB over a total range of 45–60 dB). An analogous contrast is found in the data of the present study: The large intersubject differences in masked thresholds shown in Fig. 2 are more "horizontal" than "vertical" (in terms of the axes of Fig. 2); i.e., they are mainly due to differences in phase offset, not to differences in average growth of masking. This was revealed by the analysis in Sec. I B (Figs. 3 and 5).

II. EXPERIMENT 2: PHASE EFFECT FOR A LARGE NUMBER OF FROZEN-NOISE SAMPLES

A. Rationale

All data of experiment 1 were collected using one particular 500-ms noise sample. Such an approach permits a detailed measurement and data analysis, but has the disadvantage that the data are somewhat "anecdotal;" the selected noise sample might be unrepresentative in some unknown sense, and might not reflect the masking properties of the ensemble from which it is drawn. The second experiment was designed to test this possibility by using a large set of different frozen-noise maskers.

Simply repeating the measurement scheme of experiment 1 with a lot of different frozen-noise samples would be very time consuming (measuring thresholds at eight different signal phase values for each of the samples). Furthermore, it would not yield an objective estimate of the significance of the phase effect. An obvious tool for testing the significance of the phase effect for a large number of different noise samples is an analysis of variance (ANOVA). In such an approach, a set of data is collected using a number of different noise samples and, for each noise sample, a number of different phase values. In terms of an ANOVA, "noise sample" and "signal phase" are the two factors of the analysis. At first sight, the significance of the second factor, signal phase, is the desired quantity. But signal phase has no absolute meaning across different noise samples; what really matters is only the *relative phase* of the noise and the signal.

That signal phase as such cannot have a significant effect on the thresholds can be seen as follows. Assume that for a certain set of frozen-noise samples the thresholds have been measured for two signal phase values, 0 and 180 deg. Now imagine an arbitrary subset of the noise samples to be flipped in sign. This does not affect the statistics of the set of noise samples (independent Gaussian samples) and, therefore, should not affect the statistics of the data, either. On the other hand, flipping the sign of a noise sample will reverse

the roles of 0- and 180-deg signal phase. From this it is clear that there cannot be any systematic difference between 0- and 180-deg data. The same argument applies to phase values other than 0 and 180 deg.

All this does not mean that signal phase cannot have an effect on the measured thresholds; the only restriction is that there cannot be a *systematic* effect of signal phase on thresholds obtained with *different noise samples*. In other words, if there be an effect, it is expected to be different for different noise samples. In terms of an ANOVA, this dependence is expressed by the *interaction* of the two factors, signal phase and noise sample. Experiment 2 is designed to test the significance of this interaction.

In order to maximize the number of independent frozen noise samples, we chose to use different sets of noise samples for each of the subjects.[10] This approach prevents an elaborate data analysis as in experiment 1, as well as a meaningful intersubject comparison. On the other hand, the present approach yields an objective measure for the significance of the phase effect. This is most useful in determining conditions under which the phase effect does *not* occur. In order to also test this falsifying capability of the method we applied it to two experimental conditions. The first condition was similar to that used in experiment 1 (masker components *above* the signal frequency). For the second condition we used noise maskers consisting of components *below* the signal frequency. In the latter situation it is unlikely that masking is determined by combination products (Goldstein, 1967; Greenwood, 1971; Erdreich, 1977).

B. Method

The procedure was the same as that used in experiment 1 except for two modifications. First, the masker duration was decreased from 500 to 300 ms. Second, a two-interval, two-alternative forced-choice paradigm was used instead of the three-interval paradigm of experiment 1. Both modifications served to speed up the measurements so as to allow the inclusion of a larger set of noise samples.

Two masker conditions were examined. The Gaussian noise masker, presented at a level of 65 dB SPL, was band-limited between 1 and 2 kHz for the first condition and between 100 and 800 Hz for the second condition. The method of noise generation has already been described in Sec. I A 1. In each adaptive run a single (frozen) noise sample was used as a masker. The signal was a 900-Hz sinusoid with a total duration of 50 ms. It was temporally centered in the masker and provided with 10-ms Hanning ramps.

Different sets of 20 independent noise samples were used for each subject. For each noise sample, two signal conditions were examined, differing in starting phase by 180 deg. Each stimulus condition was measured twice, which resulted in a total of 320 data points for each of the two noise types (80 noise samples, two signal phases, and two repetitions). The order of presentation was completely randomized.

The four subjects of experiment 1 participated in this experiment.

C. Results and discussion

The data sets obtained for the two masker conditions were analyzed separately. A two-way analysis of variance (noise sample versus signal phase; mixed model in which signal phase was treated as a fixed factor and noise sample as a random factor) was performed on the data obtained with the high-frequency (1 to 2 kHz) masker. This ANOVA revealed a significant effect of both noise sample ($F_{79,160}=7.22$, $p<0.0001$) and the interaction of noise sample with phase ($F_{79,160}=5.63$, $p<0.0001$). The main effect of phase was not significant ($F_{1,160}=0.88$, $p>0.05$). The insignificance of phase by itself was expected, since signal phase has no absolute meaning across noise samples. The influence of the *relative* phase of signal and masker is expressed by the interaction of noise sample with signal phase. The significance of this interaction shows that the phase effect observed in experiment 1 is not an artifact of the particular noise sample used as a masker.[11] The mean threshold was 34.7 dB; the root-mean-square error ("measurement error") was 2.3 dB.

The significance of noise sample agrees with the observations of the pilot experiment (Sec. I A 3) and also with the variability due to differences between individual noise samples reported in studies in which an on-frequency masker was used (e.g., Hanna and Robinson, 1985).

The same analysis of variance, performed for the data obtained with the low-frequency (100–800 Hz) masker, showed a significant effect of noise sample ($F_{79,160}=10.58$; $p<0.0001$). Phase and its interaction with noise sample were not significant ($F_{1,160}=0.92$, $p>0.05$ and $F_{79,160}=1.20$, $p>0.05$, respectively). The mean threshold was 37.9 dB; the root-mean-square error was 2.1 dB.

The data of experiment 2 permit a direct comparison between the two masker conditions. The observation that signal phase does not affect masked thresholds in the case of a low-frequency masker indicates a basic, qualitative difference between masking by the high- and the low-frequency masker. The two masker conditions have in common that the frequency separation between signal and masker is 100 Hz. It was pointed out in Sec. I C 1 that, in view of the 50-ms signal duration, phase-dependent signal–masker interactions are highly unlikely to occur under these conditions. As these arguments are equally valid for both masker conditions, the differences suggest another origin of the phase effect.

From studies of combination *tones* it is known that distortion products at frequencies below the primary frequency are perceptually more prominent than distortion products at frequencies above the primary frequencies. The latter are generally masked by the primaries (Zwicker, 1955), and only in special cases can they be detected by means of an indirect method (Erdreich, 1977). Thus the observed contrast between the high- and the low-frequency masker presents further support for an explanation of the phase effect in terms of auditory nonlinearities, in addition to the arguments presented in Sec. I C.

III. SUMMARY AND CONCLUSIONS

In experiment 1 an effect of signal starting phase on masked thresholds of a tone in frozen noise was found for a

condition with spectrally nonoverlapping masker and signal. The threshold-versus-phase patterns were similar to patterns obtained in other studies, in which the signal was placed spectrally within a frozen-noise masker. The occurrence of the phase effect for a nonoverlapping stimulus condition indicates that the masked thresholds were influenced by distortion products. The most probable explanation is that the signal was actually masked by odd-order distortion products below the low cutoff frequency of the masker noise band. This mechanism had previously been proposed by Greenwood (1971), Zwicker and Bubel (1977), and other investigators for comparable spectral conditions. This explanation is supported by other aspects of the data, such as the downward phase shift of the patterns with increasing masker level.

In experiment 2 many different noise samples were used as a masker in a similar frozen-noise paradigm. For each noise sample only two signal conditions were examined, differing in starting phase by 180 deg. Two masker types were used, a high-frequency and a low-frequency masker. A statistical analysis indicated that signal phase affected masked thresholds only in the case of the high-frequency masker.

The results of these experiments present direct evidence for the influence of auditory distortion products of a single band of noise on its masking potency toward frequencies below its spectrum. The experimental methods of the present study can be used to examine this influence in more detail, by systematically changing the stimulus parameters. They may supplement the insight into auditory nonlinearities gained in the study of combination tones.

ACKNOWLEDGMENTS

[1]The interpretation of Lutfi's (1983) data is somewhat obscured by his use of multiplied noise with only one noise source. As a result of this procedure, the two primary noise bands were perfectly comodulated, which might have affected the generation of the combination bands (see Sec. I C 1).

[2]An exception is the combination of sinusoidal stimuli with harmonically related frequencies. Because in this case the total waveform is periodic, the phase of both primaries can be consistently defined with respect to the period of a common fundamental frequency. Indeed, phase effects have been observed in pure-tone masking using harmonically related signal and masker frequencies (e.g., Clack *et al.*, 1972).

[3]The noise-sample selection procedure described in the text is not based on any assumptions about the generation of distortion products; it equally applies to on-band situations, where nonlinearities play no role in the explanation of phase effects.

[4]Only main effects could be included in this test since only one value for the dependent variable was available for each condition (as determined by masker level and subject).

[5]Taking one of the other subjects as a reference did not lead to a perceptible difference. A mathematically rigorous method for evaluating the relative phase of all patterns in Fig. 2 would be the maximization of the 16th-order cross-correlation function of all patterns. This was beyond our numerical scope. Since all cross-correlation functions encountered in our two-step analysis were smooth and single peaked, we assume that our method was sufficient for determining the desired phase relations.

[6]The assumption that the temporary envelope decrease of a fluctuating stimulus is not detectable is consistent with Erdreich's (1977, Sec. II A) method for removing "secondary peaks" in the monaural phase effect observed in tone-on-tone masking data. The addition of extra masking noise "restores the monaural phase effect to the form of a single sinusoid." The idea is that a temporary decrease of the (flat) envelope of a tonal masker

can be detected, but that this cue is eliminated by the extra noise.

[7]The signal duration was assumed to be reciprocal in signal power at threshold, which led to a 4.5-dB correction. A further 3-dB correction was taken to account for the 1-oct difference in signal frequency, corresponding to an approximate doubling of the critical band.

[8]In this derivation the effect of frequency separation of the components is ignored. This does not affect the argument for two reasons. First, the distribution of separations of primaries will hardly be changed if a large number of closely spaced components is increased to a still larger number. Second, the small effect of increasing the number of primary components is the introduction of *smaller* distances. According to Greenwood's (1971) construction, this will only result in a further *increase* of the power contained in the aggregate.

[9]In addition, level dependences of combination tones derived from masking experiments might differ from those measured using the cancellation method as in Zwicker (1979); Greenwood (1972) reported a slightly decreasing CT-to-primary ratio with an increasing primary level.

[10]In this respect, the four subjects are treated as one single subject; a justification for this strategy might be found in the high degree of similarity between the data of the different subjects obtained in the first experiment.

[11]Strictly speaking, the confound of noise sample and subject in the design of Exp. 2 does not exclude that the significant interaction between phase and noise sample is (partly) due to a phase–subject interaction. In view of the physical irrelevance of signal phase by itself, this is highly unlikely. Nevertheless, we ruled out the confound by performing separate two-way analyses (phase versus noise sample) on the data of each of the subjects. For all four subjects, the interaction of phase and noise sample was found to be significant: $F_{19,40} = 3.08$, 4.06, 4.49, and 9.36 for subjects A–D, respectively. These F values correspond to $p < 0.002$ for subject A and $p < 0.0001$ for subjects B–D. These results bear out the conclusions regarding the significance of relative phase of signal and noise samples formulated in Sec. II C.

Boer, E. de (**1976**). "Spectral transformations by an infinite clipper," J. Acoust. Soc. Am. **60**, 960–963.

Buus, S. (**1985**). "Release from masking caused by envelope fluctuations," J. Acoust. Soc. Am. **78**, 1958–1965.

Clack, T. D., Erdreich, J., and Knighton, R. W. (**1972**). "Aural harmonics: The monaural phase effects at 1500 Hz, 2000 Hz, and 2500 Hz observed in tone-on-tone masking when $f_1 = 1000$ Hz," J. Acoust. Soc. Am. **52**, 536–541.

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

Erdreich, J. (**1977**). "Intermodulation product pair $2f_l \pm f_h$: Masking and growth," J. Acoust. Soc. Am. **62**, 1252–1255.

Gilkey, R. H., Robinson, D. E., and Hanna, T. E. (**1985**). "Effects of masker waveform and signal-to-masker phase relation on diotic and dichotic masking by reproducible noise," J. Acoust. Soc. Am. **78**, 1207–1219.

Glasberg, B. R., and Moore, B. C. J. (**1994**). "Growth-of-masking functions for several types of maskers," J. Acoust. Soc. Am. **96**, 134–144.

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

Goldstein, J. L. (**1970**). "Aural combination tones," in *Frequency analysis and periodicity detection in hearing*, edited by R. Plomp and G. F. Smoorenburg (Sijthoff, Leiden), pp. 230–247.

Greenwood, D. D. (**1961**). "Auditory masking and the critical band," J. Acoust. Soc. Am. **33**, 484–502.

Greenwood, D. D. (**1971**). "Aural combination tones and auditory masking," J. Acoust. Soc. Am. **50**, 502–543.

Greenwood, D. D. (**1972**). "Masking by combination bands: estimation of the levels of the combination bands $(n+1)f_l - nf_h$," J. Acoust. Soc. Am. **52**, 1144–1154.

Hanna, T. E., and Robinson, D. E. (**1985**). "Phase effects for a sine wave masked by reproducible noise," J. Acoust. Soc. Am. **77**, 1129–1140.

Hawkins, J. E., and Stevens, S. S. (**1950**). "The masking of pure tones and of speech by white noise," J. Acoust. Soc. Am. **22**, 6–13.

Heijden, M. van der, and Kohlrausch, A. (**1995**). "The role of envelope fluctuations in spectral masking," J. Acoust. Soc. Am. **97**, 1800–1807.

Klitzing, R. von, and Kohlrausch, A. (**1994**). "Effect of masker level on overshoot in running- and frozen-noise maskers," J. Acoust. Soc. Am. **95**, 2192–2201.

Langhans, A., and Kohlrausch, A. (**1992**). "Differences in auditory performance between monaural and diotic conditions. I: Masked thresholds in frozen noise," J. Acoust. Soc. Am. **91**, 3456–3470.

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

Lutfi, R. A. (**1983**). ''Simultaneous masking and unmasking with bandlimited noise,'' J. Acoust. Soc. Am. **73**, 899–905.

Lutfi, R. A., and Yost, W. A. (**1981**). ''Measurement of the $2 f1 - f2$ cubic difference tone with the binaural masking-level difference,'' J. Acoust. Soc. Am. **69**, 216–222.

Mott, J. B., and Feth, L. L. (**1986**). ''Effects of the temporal properties of a masker upon simultaneous-masking patterns,'' in *Auditory Frequency Selectivity*, edited by B. C. J. Moore and R. D. Patterson (Plenum, New York), pp. 381–386.

Pfafflin, S. M., and Mathews, M. V. (**1966**). ''Detection of auditory signals in reproducible noise,'' J. Acoust. Soc. Am. **39**, 340–345.

Smoorenburg, G. F. (**1972**). ''Combination tones and their origin,'' J. Acoust. Soc. Am. **52**, 615–632.

Zwicker, E. (**1955**). ''Der ungewöhnliche Amplitudengang der nichtlinearen Verzerrungen des Ohres,'' Acustica **5**, 67–74.

Zwicker, E. (**1979**). ''Zur Nichtlinearität ungerader Ordnung des Gehörs,'' Acustica **42**, 149–157.

Zwicker, E., and Bubel, G. (**1977**). ''Einfluβ nichtlinearer Effekte auf die Frequenzselektivität des Gehörs,'' Acustica **38**, 67–71.