A comparison of masking by tones and noise

DOI:
10.6100/IR435554

Document status and date:
Published: 01/01/1995

Document Version:
Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

A comparison of masking by tones and noise

Proefschrift

ter verkrijging van de graad van doctor aan de Technische Universiteit Eindhoven, op gezag van de Rector Magnificus, prof.dr. J.H. van Lint, voor een commissie aangewezen door het College van Dekanen in het openbaar te verdedigen op vrijdag 21 april 1995 om 14.00 uur

door

Marcel Lodewijk van der Heijden

geboren te Tilburg

Dit proefschrift is goedgekeurd door de promotoren

prof. dr. A.J.M. Houtsma
prof. dr. C. Trahiotis

en de copromotor

dr. A. Kohlrausch

Omslagontwerp: Janine Mariën.

A comparison of masking by tones and noise

Contents

Chapter 1

General introduction

An historically important problem in the scientific literature concerning audition is the specification of the factors which limit listener's abilities to detect desired sounds in the presence of extraneous or unwanted sounds. The general finding is that the extraneous sounds make it more difficult to detect the desired sounds, an outcome termed 'auditory masking'. It has long been thought that the detectability of a particular sound in the presence of a second sound primarily depends on the degree to which the two sounds overlap both temporally and spectrally (*i.e.*, contain the same frequency components). The term 'critical band' is used to indicate the limited band of frequencies that affect the audibility of a signal of a certain frequency (Fletcher, 1940). The critical band concept has been used for several decades both to describe which sounds obscure others and to explain the basis of the masking effect.

The critical band concept by itself is not sufficient to explain masking phenomena. It is also necessary to specify how the information within the critical band is utilized by the listener. Many fundamental data can be summarized by noting that the listener requires a particular signal-to-noise power ratio in order to be able to detect the presence of the signal reliably. Although this approach was quite successful, it was modified and improved upon by Zwicker and his colleagues (*e.g.*, Zwicker 1970) who changed the focus from the physical relation of signals to maskers to their respective internal representations as patterns of activity in the auditory periphery. In the recent past, this type of approach has gained popularity both because of its overall success and because digital computers can now be easily used to build mathematical models of the transformations thought to occur in the auditory periphery (*e.g.*, Glasberg and Moore, 1990).

Models in which masking is explained by a combination of peripheral filtering and signal-to-noise power ratio are called 'power-spectrum models' or 'excitation-pattern models' of masking. The recent popularity of such peripherally based models, however, should not overshadow several historically important masking data which are not easily incorporated into such a framework and, therefore, restrict the generality of power-spectrum models of masking. The difficulties posed by such data are of two sorts: 1) inherent nonlinearities produced by processing within the auditory periphery and 2) the findings that audibility is sometimes determined by factors other than the signal-to-masker power ratio. Details concerning these two factors which constrain applicability of the power-spectrum model are provided below.

The major contributions of this thesis are 1) an evaluation of under what conditions and how well masking data can be accounted for by computational models based on peripheral excitation-patterns and 2) an experimental investigation providing further data which militate against the validity and applicability of excitation pattern based models of masking.

1.1 Frequency selectivity and auditory nonlinearity

The frequency selectivity of the ear has been the starting point of many attempts at modeling auditory phenomena. Many of these investigations deal with pitch perception rather than with masking. For example, Helmholtz (1885) sought to explain pitch perception in terms of a spectral analysis performed by the inner ear. His approach focused on the physiological response of the inner ear to sounds. Such a liaison of psychoacoustics with physiology has since been the focus of many investigations (*e.g.*, Fletcher 1930, Greenwood 1961b). It was stimulated by the availability of increasingly sophisticated techniques for studying (electro-)mechanical and neural aspects of the hearing organ. Physiological studies of the auditory periphery have made clear that a substantial frequency analysis takes place at the level of the cochlea. For an overview, the reader is referred to Pickles (1988).

The concept of the critical band, arising from Fletcher's (1940) study, had a large impact on both theoretical and experimental work in audition. It has been used to explain a variety of auditory phenomena including frequency discrimination (Zwicker, 1970), loudness of complex sounds (Scharf, 1970) and masking itself. The observation that only sounds in a limited band of frequencies affect the audibility of a tone suggests the involvement of some filter-like process. An obvious way of modeling auditory filtering is by analogy with resonance characteristics frequently encountered in physics, e.g., those of an electrical tuning circuit. This approach can no better be illustrated than by the following quotation from Schafer et al. (1950, pp. 491):

> In this paper the selectivity characteristic of the ear is interpreted in terms of a simple "filter," a filter being a familiar device having frequency selective properties which are useful for discriminating signal from noise. This interpretation is for analytical purposes only, and is not intended to suggest the mechanism by which the selectivity is obtained. The fact that the mechanism is a complex interrelation of peripheral elements of hydrodynamical and mechanical nature, neural elements electrochemical in nature, and central neural processes discourages attempts at carrying the analogy further.

A difficult challenge for this kind of modeling of auditory frequency selectivity is provided by nonlinear aspects of the hearing system. For example, the micromechanical properties of the mammalian cochlea are known to be highly nonlinear (Pickles, 1988). Likewise, inherent nonlinearities of the auditory system have been discovered in many different psychoacoustical experiments. In such experiments, nonlinearities are revealed whenever the phenomena being observed do not display proportionality or superposition. An example is the formation of distortion products in the ear, *i.e.*, components at frequencies not present in the physical stimulus. When the stimulus consists of a single tone, harmonic distortion can lead to so-called aural harmonics. When two or more primaries are presented to the ear, distortion within the ear can lead to inter-modulation products called combination tones. The effects of both types of distortion products on masking data is briefly discussed below.

An early indication that aural harmonics affect masking was provided by Wegel and Lane (1924). They measured tone-on-tone masking as a function of the level of the masker. They found that their patterns of masking were neither single-peaked, nor smooth, as would be expected if only sound within the critical band determined performance. Their masked thresholds showed a sharp drop in the immediate vicinity of the masker frequency. For those cases, the thresholds were determined by the detection of beats produced a direct interaction between masker and signal waveforms. Similar notches occurred at harmonic frequencies of the masker, particularly for high masker intensities. This indicated that the listeners were perceiving beats between the signal and higher harmonics of the masker. The higher harmonics of the masker were called 'aural harmonics' by Wegel and Lane to emphasize that they originated from the auditory system. That is, the beats were not due to distortions generated by the apparatus.

Secondary peaks in tone-on-tone masking patterns have also been observed at non-harmonic frequencies of the masker frequency (e.g., Ehmer, 1959). As an analysis by Greenwood (1971) has shown, these irregularities stem from the generation of combination tones in the ear. The explanation is based on the fact that, in a limited region above the masker frequency, the nonlinear interaction between signal and masker generates distortion products *below* the masker frequency. As a result of the spectral asymmetry of masking, these combination tones are often more audible than the signal itself. The level of the combination tones depends strongly on the separation between the primaries; when the signal frequency is increased sufficiently, the combination tones become too weak to be detected. The secondary peak in the masking pattern is caused by this breakdown of the extra cue with increasing masker-signal separation. Among the combination tones, the odd-order tones at frequencies $(n+1)f_l - nf_h$ are most salient (f_h and f_l are the high and low primary frequencies, respectively; n is a small positive integer). In contrast to harmonic distortion and simple difference tones ($f_h - f_l$), the occurrence of odd-order products is not restricted to high primary levels (Zwicker, 1955; Goldstein, 1967).

Due to the large effect of aural harmonics and combination tones on measured masked thresholds, the 'true' auditory frequency selectivity can be obscured by spectral components which are not present in the acoustic stimulus. Although some types of distortion products occur over a wide dynamic range, it is not unreasonable to consider distortion products an 'artifact' and to insist on 'uncorrupted' data for the estimation of the ear's frequency selectivity. This can usually be achieved by avoiding high stimulus intensities and/or by the addition of background noise to mask combination tones (Greenwood, 1971). But even when such precautions are taken, many masking data show nonlinear aspects of the auditory system. An important example is the nonlinear growth of masking in a situation where the masker and the signal are spectrally distant.

A considerable part of Wegel and Lane's (1924) analysis of tone-on-tone masking is concerned with this nonlinear growth of masking, revealed by non-unity slopes (on a log-log scale) of the curves relating signal level at threshold to masker level. Slopes larger than unity (expansive or more-than-proportional growth of masking) were observed for signal frequencies above the masker frequency. The slopes for this spectral condition can be as high as 3 dB/dB (Schöne, 1977). Wegel and Lane attributed this so-called upward spread of masking to aural harmonics of the masker, motivated by the fact that nth order harmonic distortion products grow as the nth power of the primary level. Later studies, however, revealed that the upward spread is not due to harmonic distortion (Egan and Klumpp, 1951; Ehmer, 1959). Masking towards lower frequencies shows the opposite tendency: the slopes of the growth-of-masking curves are less than unity; slopes as low as 0.3 dB/dB have been observed in tone-on-tone masking (Schöne, 1977).

The highly nonlinear growth of masking is a serious problem for the description of the auditory periphery in terms of filters. Attempts to cure excitation-pattern models of their too linear behavior with the help of parametric, level-induced bandwidth changes (Glasberg and Moore, 1990) run the risk of inconsistencies or ambiguities (Rosen and Baker, 1994). It may be easier to model auditory frequency selectivity than it is to model auditory nonlinearity, but the latter is anything but a secondary aspect of hearing.

1.2 Signal-to-noise ratio as criterion of detectability

It is not a trivial problem how to relate behaviorally measured masking patterns to assumed internal excitation patterns of the inner ear evoked by the masker. An early attempt at quantifying the relation between these two types of patterns can be found in Fletcher's "space-time pattern theory of hearing" (1930, p. 328):

> If we assume that the masked tone is just perceptible when it is at a level corresponding to that necessary when the masking and masked tone have the same frequency then the data on minimum perceptible differences in intensity can be applied.

This procedure is a reduction of masked thresholds to intensity difference limens. In the case of identical masker and signal, this reduction is self-evident. Or as Miller (1947, p. 615) puts it: "when the masked and masking sounds are identical, the difference between masking and sensitivity to changes in intensity lies only in the way the story is told." Fletcher's assumption, however, is a non-trivial generalization of this equivalence, enabling him to calculate the "pattern of stimulation of the inner ear," starting from the masking pattern obtained in a psychoacoustic experiment. Two modern attempts to relate masking and intensity discrimination have been provided by Bilger (1976), who related several studies in the literature and Wier *et al.* (1984), who related relevant data collected from one set of subjects.

In many subsequent models of masking (Fletcher, 1940; Schafer *et al.*, 1950; Patterson, 1976; Moore and Glasberg, 1987b), the identification of masked thresholds with intensity difference limens is further simplified by assuming that the audibility of a signal requires a particular signal-to-masker power ratio.[1] The assumption of signal-to-masker power ratio as the detection criterion, however, is at odds with a large body of psychoacoustical data.

First, it fails to account for the fact that detectability improves with signal duration (*e.g.*, Plomp and Bouman, 1959). The data clearly indicate that power integrated over time (energy) determines detectability. In addition, when the masker is not deterministic, the data can be accounted for by an 'energy-detection model' that takes into account the durations and bandwidths of the stimuli (Green and Swets, 1966). This contrasts with most power-ratio models that neglect the stochastic nature of the stimulus.

A second, classical, example showing that detection is not simply based on signal-to-noise ratio is the listener's use of beats in pure-tone masking experiments (*e.g.*, Wegel and Lane, 1924). Those experiments indicate that detection can be based on fluctuations produced by adding a tonal signal of one frequency to a tonal masker of another frequency.

Recently, dozens of experiments have investigated listeners' use of spectral shape to detect signals. Detection is accomplished by comparisons of spectra across critical bands or 'frequency channels'. This type of processing of stimuli is called 'profile analysis' and many important data are summarized in Green's (1988) monograph. Virtually all studies of profile analysis include randomly determined, trial-by-trial, and observation-interval-by-observation-interval changes in the overall level of the stimuli. This 'roving-level paradigm' precludes the listener's use of energy-based cues. An especially interesting experiment was published by Kidd *et al.* (1989) who

[1] This simplification is correct if relative intensity difference limens are independent of level (Weber's law), which is only approximately true.

showed that the detectability of a tone in a narrow-band noise was not severely degraded by roving the levels of the stimuli.

Ironically, it was the detectability of tones in narrow bands of noise that led Fletcher (1940) to formulate his critical band concept. Fletcher's interpretations of the data focused on the recurring constant values of signal-to-noise ratio at threshold and omitted descriptions of the perceptual nature of the cues used by the listeners. Explicit examples of alternative cues used by listeners can be found in Buus (1985) and Richards (1992).

Perhaps the greatest challenge to the power-ratio view has been provided by Hall and his colleagues (Hall *et al.*, 1984; Hall and Grose, 1990). They have shown that detectability of a tone masked by a narrow band of noise can be improved when a second band of noise, at a different center frequency, is added. A necessary condition is that the 'off-frequency band' of noise and the masker have identical or 'co-modulated' envelopes. Such influence of off-frequency information clearly violates the notion that the detectability is solely based on the signal-to-noise ratio in the critical band. All of these examples show that the power-ratio view is incomplete and also probably indicate that a monistic view of masking is inconsistent with the wealth of experimental facts.

1.3 Outline of this thesis

This thesis focuses on those instances of masking which can be described as spectral or stationary masking and for which, at a first glance, a straightforward explanation in terms of auditory frequency selectivity is possible. To put it naively, those situations are considered for which the power spectrum model should do a good job - after all it has been developed to cover exactly these conditions. By the 'power spectrum model of masking' is meant, not just one particular model but, rather, all of the assumptions and views described previously. The main question is: does such a model really describe these situations adequately and if not, what is missing or wrong?

In Chapter 2, a quantitative excitation-pattern model proposed by Glasberg and Moore (1990) is tested with respect to its usability in predicting masking. This is done in two steps. First, it is analyzed how such a model should be used to simulate masking experiments. It is argued that the way in which the model was derived from experimental data is in fact reversible, so that really no new assumptions are needed to use the model backwards, *i.e.*, as a predictor of masking. The assumptions on which the model is based suggest a wider range of applicability than the specific type of experiments which form the quantitative basis of the model. The second part of Chapter 2 contains a test of this generalizability. Various sets of psychoacoustical data, reported in the literature are compared with predictions of the model. The discussion of the results of this comparison concentrates on systematic deviations rather than on incidental, minor mismatches.

The main weak point of the model appears to be that it fails to reproduce the difference in masking by tones and noise bands that is clearly present in the experimental data. If the masker is a tone and the signal is spectrally distant from it, the model behaves far too linearly. Given this failure, the original question 'what is wrong with the model' is transformed to: what makes tones and bands of noise behave differently as maskers? The analysis of Chapter 2 only reveals that this difference is not covered by the power-spectrum model of masking. In order to obtain more concrete information, a number of psychoacoustic experiments on specific aspects of noise maskers were conducted. Chapters 3 and 4 report on this experimental work. Chapter 3 deals with the role of masker envelope fluctuations for masking towards higher frequencies. In Chapter 4, the role of aural distortion products is studied for maskers consisting of a single band of noise. Chapter 5 summarizes the main conclusions.

Chapter 2

Using an excitation-pattern model to predict auditory masking[1]

...we must admit that we have not yet reduced psychoacoustics to symbols.
J. C. R. Licklider in 1951

Abstract

This Chapter evaluates the extent to which auditory masking can be predicted from excitation patterns. For this purpose a quantitative model proposed by Glasberg and Moore [Hear. Res. 47, 103–138 (1990)] was used to calculate excitation patterns evoked by stationary sounds. Model simulations were performed for a number of masking experiments, reported in the literature, by calculating excitation patterns for the masker-alone as well as for the masker-plus-target conditions. As a threshold criterion, a difference between the two patterns of 1 dB at any frequency was imposed. For narrow-band-noise masking patterns, the method yields a fairly precise prediction of experimental data. For other conditions, however, systematic deviations between model predictions and data are observed. For instance, the model does not reproduce the typical tip-tail shape of psychoacoustical tuning curves. Furthermore, the nonlinearities in level dependence are not correctly described, and the model fails to reproduce a realistic two-tone masking curve.

[1]Based on a paper, written together with Armin Kohlrausch, which appeared in *Hearing Research* 80 (1994), 38–52.

2.1 Introduction

Spectral aspects of masking are generally believed to reflect the spread of excitation along the tonotopically organized sensory structures in the auditory periphery. Models to calculate excitation patterns evoked by an auditory stimulus have been proposed by several authors in order to explain or predict certain aspects of hearing related to spectral resolution.

Helmholtz (1885, p. 146), in his discussion on the theory of the function of the cochlea, provided an example of what nowadays would be called an excitation pattern. Referring to a resonance curve, he wrote: "...(this Figure) might be taken to represent .. a longitudinal section of that part of the basilar membrane in which the proper tone of the radial fibers of the membrane are nearest to the exciting tone". This view is part of his 'place theory' of pitch.

Zwicker (1956, 1970) presented a quantitative model to calculate excitation patterns from tone-on-tone masking data. In fact, his excitation patterns can be characterized as 'normalized versions' of measured masking patterns. Zwicker used his patterns to explain aspects of loudness, intensity discrimination and frequency discrimination. Florentine and Buus (1981) studied a multiband version of Zwicker's (1956) and Maiwald's (1967) excitation-pattern model for intensity discrimination.

A different approach is related to the power-spectrum model of masking (Fletcher, 1940; Patterson, 1974). In this model the auditory periphery is conceived as containing a bank of bandpass filters. Masking is then determined by the signal-to-noise ratio at the output of the filters. Since the masker is not necessarily noisy, it is more appropriate to speak about target-to-masker (T/M) ratio. The signal is masked if this ratio does not exceed a certain critical value. Under these assumptions one can estimate filter characteristics from masking data. These characteristics include filter bandwidth and shape. The excitation pattern is identified with the pattern of response in the filter channels as a function of characteristic frequency; knowledge of the filter characteristics makes it possible to calculate the excitation pattern evoked by an arbitrary stimulus.

When predicting thresholds of tones in the presence of a *broadband* noise, only the width of the auditory filter is important, not its exact shape (Fletcher, 1940). In this situation the filters can be modeled to have a rectangular shape. Under this assumption, the width of a filter (the critical bandwidth or its equivalent rectangular bandwidth, ERB) can be derived in a straightforward manner from broadband noise-on-tone masking data, apart from an overall scaling factor corresponding to the efficiency of the detection.

Patterson (1974) used masking of tonal targets by low pass and high-pass noise to calculate filter shapes at various frequencies. Subsequent refinement of experimental and calculational methods led to a model for calculating auditory-filter shapes from notched-noise masking data, proposed by Patterson (1976), Moore and Glasberg (1987b) and Glasberg and Moore (1990). With the help of these filter shapes, excitation patterns of arbitrary stimuli can be calculated. Moore and Glasberg suggest their use in evaluating different aspects of hearing, including masking of arbitrary stationary maskers (Moore and Glasberg 1987b, p. 221). Their model is described in more detail in the next section.

Of course, the derivation of filter shapes from masking data, mentioned above, can only be correct if at least the following underlying assumptions are realistic:

- masking directly reflects peripheral filtering;

- peripheral filtering is adequately described by a bank of bandpass filters;

- threshold correlates with T/M ratio at the auditory filter outputs.

In this study we test the validity of these assumptions by predicting masking behavior from the filter shapes. This procedure is applied to conditions, different from the notched-noise masking which yielded the filter shapes. These predictions are compared with psychoacoustic data from the literature, and possible differences are discussed.

The investigation of a model to predict masking is not only of academic but also of technical interest. Masking phenomena play an important role in digital coding schemes, which always produce quantization noise. For high-quality coders, this noise has to be inaudible. A reliable algorithm to predict how well this noise is masked by the coded signal is therefore an essential element for the development of efficient coding algorithms (e.g., Veldhuis et al., 1989), and of so-called objective quality measures (e.g., Beerends and Stemerdink, 1992; Brandenburg and Sporer, 1992).

2.2 General method

2.2.1 Model

In order to evaluate the predictive power of an excitation-pattern model with regard to masking, we made use of a well documented model of peripheral auditory filtering, described by Glasberg and Moore (1990). Under the assumptions of the power-spectrum model of masking, they derive the filter characteristics of the auditory system from results of notched-noise masking experiments. In these experiments noise with a spectral gap is used to mask a tonal target spectrally centered in the gap. The resulting filter shapes are characterized as rounded exponential (RoEx) filters (Patterson and Nimmo-Smith, 1980). Glasberg and Moore present typical filter parameters for normal-hearing subjects. These parameters depend on the center frequency of the filter and on the stimulus level. We have digitally implemented a slightly modified version of this model, which will now be described briefly. For details, the reader is referred to Glasberg and Moore (1990).

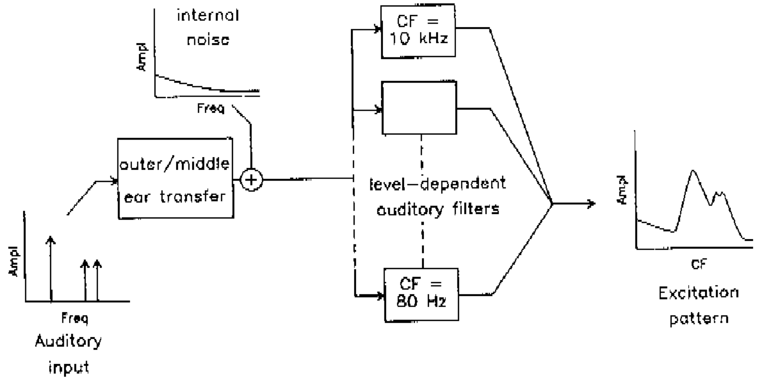

Figure 2.1: Schematic diagram of the calculation of excitation patterns. After a spectral weighting to simulate outer and middle ear transfer, the amplitude spectrum of the signal is passed through a bank of overlapping band-pass filters. The filter widths are determined by a spectral weight of the signal level. Internal noise is added to provide a lower excitation limit. The output levels of the filters as a function of their center frequency (ranging from 80 Hz to 10 kHz) result in the excitation pattern.

Fig. 2.1 shows the building blocks of the model. Excitation patterns are calculated from the amplitudes of the discrete stimulus components. The first stage of the calculation consists of a spectral weighting of the spectral components in order to simulate the transfer of the outer and middle ear. These transmission characteristics are derived from 100-phon equal-loudness curves as suggested by Glasberg and Moore (1990).

The peripheral auditory system is conceived as containing a bank of bandpass filters, which are assumed to be approximately linear. The selectivity of the filters depends on both center frequency and stimulus level. With increasing center frequency, both lower and upper filter slopes become shallower, yielding a larger bandwidth.

The only nonlinear feature is a parametric dependence of the filter slopes on the stimulus level. With increasing input level the lower slope of the filter becomes shallower. The contribution of each stimulus component to the excitation pattern is calculated with a filter shape that is particular to that component. The lower slope of a filter is determined by the total stimulus level within the one-ERB band surrounding the stimulus component under consideration[2].

The filter bank splits the stimulus up in an array of channels, each of which has its characteristic frequency (corresponding to the center frequency of the filter) in the range from 80 Hz to 10 kHz[3]. The excitation at a certain frequency is defined as the level at the output of the filter centered on that frequency.

In order to simulate the absolute threshold of the auditory system, a lower limit is set to the excitation pattern by adding 'internal noise' to the stimulus. The spectral content of this noise is chosen so as to reproduce the absolute threshold for tonal stimuli. In this way the absolute threshold is modeled as a special case of masking (see next section). The addition of internal noise is the only modification of the original algorithm of Glasberg and Moore for calculating excitation patterns.

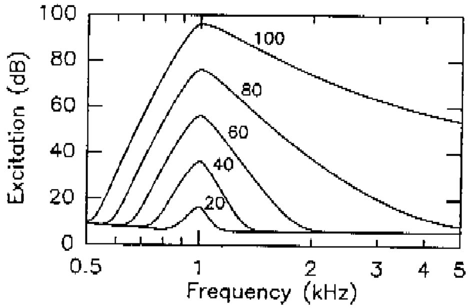

Figure 2.2: Excitation patterns evoked by a single 1-kHz tone at a level of 20, 40, 60, 80 and 100 dB SPL.

In Fig. 2.2 excitation patterns evoked by a single 1-kHz sinusoid at various levels are shown. The expansive increase of excitation at the *upper* skirts of the pattern is caused by the flattening of the *lower* filter skirts with increasing stimulus level. This upward spread of excitation correlates to the nonlinear (expansive) growth of masking with level, when a tone masks a second tone with a higher frequency (Schöne 1979; see also Fig. 2.10 below). Apart from the limitation at lower levels

[2]In the text of Glasberg and Moore (1990) it is not stated explicitly in which way the level of the input components influences the slope of a particular auditory filter. A detailed analysis of the source code of the program included in their article reveals the procedure that is used in the present paper.

[3]In our implementation the filters are equally spaced on an ERB-rate scale (see Glasberg and Moore, 1990). The density is 10 filters/ERB.

caused by the internal noise, peaks and lower skirts of the patterns grow linearly with level.

2.2.2 Procedure

Given the procedure to calculate the excitation pattern evoked by a spectrally-represented stimulus, we will now describe how we calculate masked thresholds from these patterns. In a masking experiment the subject has to discriminate between a stimulus containing the masker alone (M) and a stimulus containing both masker and target (M+T). In terms of the power-spectrum model of masking, this means that the excitation patterns evoked by these two stimuli have to differ by some critical amount at, at least, one frequency. Motivated by commonly reported intensity difference limens, we take this critical value to be 1 dB[4]. It is important to note that the frequency of maximum difference between the two patterns is *not* necessarily equal to the target frequency. In many situations the largest difference between M and M+T patterns occurs at a frequency different from the target frequency. This phenomenon is referred to as off-frequency listening (OFL)[5]. A more complete quantitative description of OFL is given in the next section.

In a simulation of a masking experiment, an iterative procedure is used in which the relative levels of masker and target are adjusted to only just meet the criterion of discrimination between M and M+T patterns. A simple example of tone-on-tone masking is illustrated in Fig. 2.3. The solid line is the excitation pattern evoked by a 1-kHz tone at a sound pressure level (SPL) of 60 dB (the masker). The dotted line is the pattern arising from a 1.3-kHz tone with a sound pressure level of 30.5 dB (the target). Its level has been calculated to just yield a 3-dB difference between the combined pattern of masker + target and the pattern of the masker alone. In this single case, a criterion of 3 dB instead of 1 dB has been used for visual clarity. The M+T pattern is given by the dashed line. The arrow indicates the point of maximum difference between M and M+T patterns, which is located slightly above the target frequency. Thus, even in this simple situation, the above-mentioned phenomenon of OFL appears to play a role in the simulation.

It should be stressed that usage of an excitation-pattern model is limited to cases in which temporal aspects of the stimuli do not dominate the masking behavior, since all stimuli are represented in the model exclusively by their spectral amplitudes. This precludes influence of beats, signal phase, onset and offset effects, envelope

[4]In most situations, taking a different criterion only leads to a uniform vertical shift of the predicted masking pattern. The criterion directly corresponds to the T/M ratio at threshold, i.e. the efficiency factor K in the derivation of the filter shape from notched-noise masking data (Glasberg and Moore 1990, Eq. 1).

[5]The term off-frequency listening was introduced by Leshowitz and Wightman (1971) for a different phenomenon, namely the detection of 'spectral splatter' resulting from steep signal on- and offsets. The present use of the term agrees with that by Johnson-Davies and Patterson (1979) and Patterson and Nimmo-Smith (1980).

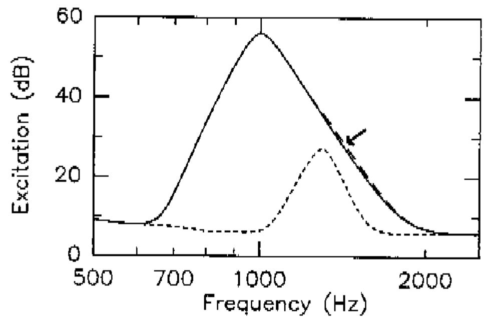

Figure 2.3: Excitation patterns illustrating the calculation of masked threshold of a 1.3-kHz target in the presence of a 1-kHz masker at 60 dB SPL. The solid line is the pattern of the masker alone (M), the dotted line indicates the target alone (T), and the dashed line is the pattern of masker and target combined (M+T). The level of the target was adjusted to only just yield a 3-dB difference between M and M+T patterns (a 3-dB criterion was applied only in this figure for the sake of visual clarity; in all simulations of masking experiments a 1-dB criterion was used). The location of the maximum difference is indicated by the arrow.

fluctuations, and other frequently encountered temporal aspects of masking. So care must be taken in the selection of masking experiments to be simulated. Only conditions that approach the ideal case of stationary masking will be considered.

The generation of distortion products by the hearing system can seriously affect masked thresholds (Goldstein, 1967). The absence of such distortion in the excitation-pattern model imposes a second limitation on its usage. Only experiments in which the occurrence of distortion products is unlikely, or in which precautions have been taken to exclude their influence, can be compared with simulations.

2.2.3 Off-frequency listening

In a simulation of a simple tone-on-tone masking condition (cf. Fig. 2.3), we noticed that detection of the target can take place at a filter which is centered at a frequency different from the target frequency. This phenomenon is called off-frequency listening (OFL). In this section we take a closer look at the occurrence of OFL and its connection with the specific shape of the auditory filter and its *quantitative* influence on predicted masked thresholds.

When a masker M is present, detectability of another stimulus component (the target T) depends on the T/M excitation ratio in the frequency range where this ratio is largest. This range is determined by the relative slopes of the excitation patterns in the region of overlap. In the Glasberg and Moore model, RoEx filters (Patterson and Nimmo-Smith, 1980) are used. The use of these filters, in combination with the formula that relates bandwidth to center frequency, results in the shape of the calculated excitation patterns. The filter slopes of RoEx filters are approximately straight when the logarithm of the attenuation is plotted on a linear Hz scale. Since filter width (ERB value) increases with center frequency, both lower and upper filter slopes become shallower with increasing center frequency. Fig. 2.4 shows an example of the excitation patterns under consideration [6]. It shows patterns

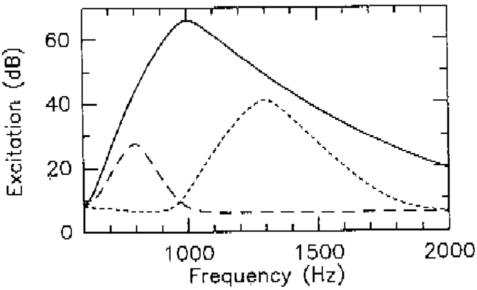

Figure 2.4: Excitation patterns of a 1-kHz masker (solid line), an 800-Hz target (dashed line), and a 1300-Hz target (dotted line). Target patterns are 'effective patterns' (see text). The sound pressure level of the masker is 70 dB; the targets have threshold levels in the presence of the masker. Important for off-frequency listening are the relative lower slopes of masker and lower target, and the relative upper slopes of masker and higher target, respectively.

of a 1-kHz masker at 70 dB (solid line), and patterns of two targets at threshold level, with frequencies of 800 (dashed line) and 1300 Hz (dotted line).

In the case of the high-frequency target, the *upper* slope of the patterns is relevant. In this region the excitation of the masker falls off more slowly than the target excitation. These different rates of decrease are a direct consequence of the flattening of the auditory filter slopes with increasing center frequency. The result

[6] Since masking is determined by the simultaneous excitation of T and M, it is necessary to adjust the filter shapes according to the over-all level of target + masker. So the excitation pattern of the target alone is in fact calculated with filters that are broadened by the presence of the masker, although the excitation of the masker itself is not included in the calculation; one could call this the 'effective' excitation of the target.

is that the T/M ratio *decreases* with increasing frequency, *i.e.*, that off-frequency listening at a higher frequency will not make the target more detectable. The point of best detectability is still slightly higher than the target frequency, but that is only a consequence of the rounded maximum of the target pattern.

In the case of the low-frequency target, the reverse effect is observed. Now the *lower* slopes of the masker and target patterns are of interest. Since filter slopes increase with decreasing center frequency, the T/M ratio *increases* with decreasing center frequency. This means that off-frequency listening will result in improved detection in this case. But, since the lower slopes of the excitation patterns are steep and unaffected by stimulus level, absolute threshold will limit the use of this cue.

To illustrate the effect of OFL quantitatively we show two calculated tone-on-tone masking patterns in the upper panel of Fig. 2.5. The masker is a 1-kHz sinusoid with a level of 70 dB SPL. The solid line is the pattern of masked thresholds calculated in the ordinary way as described in the previous section. The dashed line is calculated by restricting the detection to the target frequency, *i.e.*, without OFL. As expected from the discussion above, the largest 'release of masking' due to OFL is observed at the lower slope of the masking pattern. In the lower panel of Fig. 2.5 this masking release (threshold without OFL minus threshold with OFL) is plotted for 1-kHz maskers of 50, 70 and 90 dB SPL, respectively. For target frequencies below the masker frequency, the masking release increases with masker level because at higher levels OFL is less restricted by internal noise. For target frequencies above the masker frequency, the masking release is generally no larger then 3 dB and does not depend on the masker level.

A second illustration of the role of OFL is provided by nonlinear addition of masking. In the upper panel of Fig. 2.6 the patterns of two sinusoids are shown (800 and 1300 Hz, solid lines) each of which just masks a 1000-Hz sinusoid of 24 dB SPL (dotted line). The patterns for low masker + target and high masker + target are given by the dashed lines in Fig. 2.6. Arrows indicate the points of optimal detectability of the target in the presence of one of the two maskers. For both maskers the frequencies of optimal detection are shifted away from the masker frequency. When the two maskers are combined to form one two-tone masker (lower panel of Fig. 2.6), each of the points of best detectability in the single-masker case is spoiled by the presence of the other masker. So OFL is of no help here, and the point of optimal detection occurs approximately at the target frequency (arrow in the lower panel of Fig. 2.6). The elimination of OFL results in a more than 3-dB addition of masking; the threshold is actually increased by 11 dB on combining the maskers. Note that this 'nonlinearity' has nothing to do with nonlinearities in the calculation of the excitation patterns; in fact, the excitation at the intersection point of the two individual masker patterns increases by exactly 3 dB on combining the two maskers.

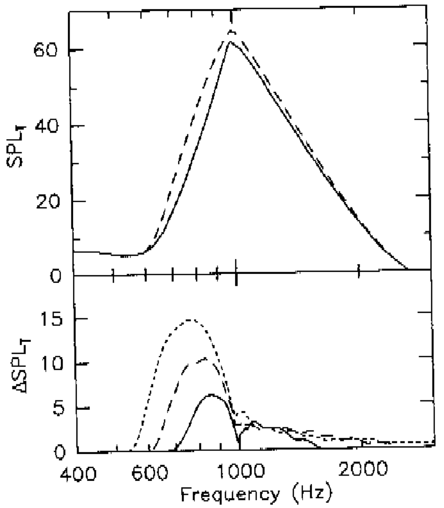

Figure 2.5: The upper panel shows two calculated masking patterns for a 1-kHz masker with a sound pressure level of 70 dB. The solid line is calculated without restrictions; the dashed line is calculated by restricting detection to the filter centered on the target frequency. The lower panel show the 'release of masking' due to off-frequency listening for a 1-kHz tonal masker with SPLs of 50 (solid line), 70 (dashed line) and 90 dB (dotted line), respectively.

2.2.4 Multi-band detection criterion

An alternative detection criterion would be one in which the information of several bands is combined or summed, namely, an integrative detection criterion (Florentine and Buus, 1981). There are several arguments against considering such an integrative detection criterion in this study.

First, the RoEx filter shapes in the model are derived without taking such integration into account. The procedure described above is really the exact reverse of the derivation of filter shapes from masking data by Glasberg and Moore (1990). By introducing new features into the model, we would not be able to judge the accuracy and degree of reliability of this model. Furthermore, for the conditions investigated in this study, we do not expect big differences by considering summation

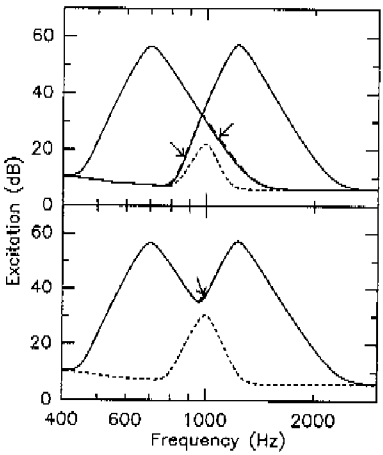

Figure 2.6: Illustration of nonlinear addition of masking due to off-frequency listening. In the upper panel, the solid lines indicate excitation patterns of two sinusoidal maskers at 700 and 1300 Hz, respectively. Each masker has a level that yields a 24 dB threshold of a 1-kHz tone (dotted line). The dashed lines are the patterns of masker + target; arrows indicate the points of best detection. The dotted line is the excitation pattern of the target at threshold alone. The lower panel shows the effect of combining the two maskers (solid line). Detection is limited to the valley between the maskers (arrow), yielding a target threshold of 35 dB (dotted line).

over different channels. For significant differences in calculated masked thresholds to occur, the T/M ratio has to be relatively constant over an appreciable spectral range. In the previous section it was shown that, in the case of the RoEx-filters used here, this condition is generally not met when using narrow-band maskers. Spectral integration is only expected to have significant effects when both masker and target are broadband signals (Gässler, 1954; Langhans and Kohlrausch, 1992a). This situation is not considered in this study, and, in fact, it is a situation in which exact filter shapes are not relevant. Finally, there is a great degree of arbitrariness in introducing spectral integration. Both the frequency- and the excitation-scale over which the 'detection variable' has to be summed, are to be chosen. This would introduce into the model a number of additional parameters that might obscure its operation and its evaluation.

The approach based on 'local detection' is as simple as possible; all simulations are carried out with the same filter parameters and detection criterion. The only free parameter is the absolute threshold of the target. When absolute thresholds are not reported in the experiments to be simulated, it is set to a default value, provided by Glasberg and Moore (1990).

In one case, *viz.*, the simulation of psychoacoustic tuning curves, we did consider the alternative procedure of integrative detection. We essentially adopted the method described by Florentine and Buus (1981) for combining the information from different channels. The reader is referred to the section on psychoacoustic tuning curves for more details.

2.3 Simulations of masking experiments

In this section, a number of psychoacoustical experiments reported in the literature are simulated by the excitation-pattern procedure described above. For each experiment, the original experimental data are compared with the model prediction. Many of the data sets considered here are backed up by additional studies which show similar trends. Our main goal is to investigate if the excitation pattern model correctly reproduces these trends. If differences between data and model predictions are encountered, we concentrate on systematic deviations, not on incidental mismatches of a few dB.

2.3.1 Masking patterns

Egan and Hake (1950) presented masking patterns of sinusoidal test signals, masked by a narrow band of noise centered at 410 Hz, with a bandwidth of 90 Hz. They used narrow-band noise (NBN) as a masker to prevent subjects from using beats between masker and target as a cue for detection.[7]

In Fig. 2.7, Egan and Hake's data, taken from their Fig. 2, are reproduced (triangles, squares and circles). Masking patterns obtained at three different masker levels (40, 60 and 80 dB SPL) are shown. The widths of the peaks give an indication of the masker bandwidth and the ear's frequency selectivity; the expansive increase of the upper slope of the pattern with level forms an example of upward spread of masking.

In the same figure, model calculations are presented by lines. The simulations reproduce the data in a fairly precise way, differences generally not exceeding 5 dB. Systematic differences are observed for target frequencies within and below the masker band, where the model predicts excessive masking. As reported by Egan

[7]In the simulations noise is spectrally represented by a number of sinusoids, too closely spaced in frequency to be resolved by the filters, e.g. 10 Hz.

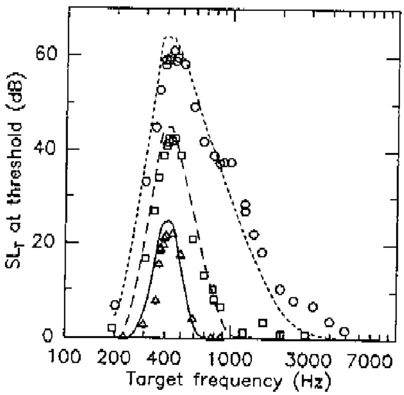

Figure 2.7: Masking patterns (symbols) of Egan and Hake (1950, Fig. 3) together with model simulations (lines). The masker is a 90-Hz-wide band of noise, centered around 410 Hz. The target is a pure tone. Three different masker levels were used: 40 (solid line, triangles), 60 (dashed line, squares), and 80 dB SPL (dotted line, circles). Target thresholds are expressed in dB SL (Sensation Level).

and Hake (1950) detection within the masker band was influenced by a form of beat detection.

Zwicker and Jaroszewski (1982) measured tone-on-tone masking patterns for a number of masker levels. The data points in Fig. 2.8 show three of these patterns obtained with a tonal masker at 1 kHz at levels of 20, 40 and 60 dB SPL. Low-pass noise with a cut-off frequency of 0.8 times the target frequency was added to mask distortion products.

The peak value in each masking pattern is obtained for the case target frequency f_T equal to the masker frequency f_M, which corresponds to an intensity discrimination task (masker and target are added with a phase difference of 90 degrees). In a narrow frequency range around the masker frequency (one or two data points on either side of the peak), beats between masker and target may have influenced the measured thresholds.

As Zwicker and Jaroszewski pointed out, increasing the masker level has the effect of reversing the asymmetry of the patterns: at low levels, masking towards lower-frequency targets is greater than towards higher-frequency targets, whereas at higher levels this effect is reversed. The model (lines in Fig. 2.8) also predicts

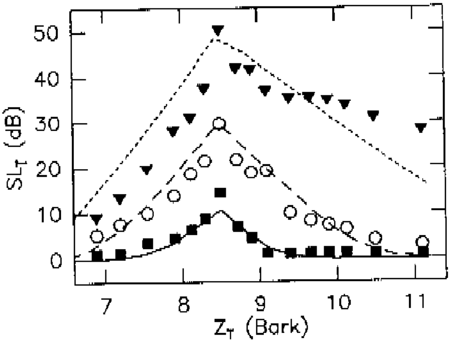

Figure 2.8: Tone-on-tone masking patterns and model predictions for three different levels. The masker frequency is 1 kHz. Data from Zwicker and Jaroszewski (1982, Fig. 2) for masker levels of 20 (squares), 40 (circles), and 60 (triangles) dB SPL. Model predictions are indicated by lines.

increased masking towards higher frequencies with increasing level (upward spread, due to increasing filter bandwidth with level), but the effect is weaker than in the experimental data. This results in an *underestimation* of masking towards higher frequencies at high masker levels. Furthermore, the model fails to describe the increase of the lower slope of the masking pattern with increasing masker level. This leads to an *overestimation* of masking towards lower frequencies at high masker levels. Summarizing, we can say that the model is behaving too linearly in this simulation of tone-on-tone masking.

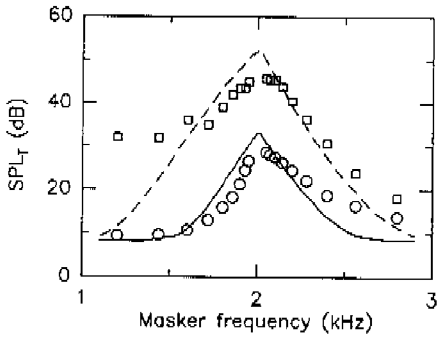

Figure 2.9: 'Mirrored' tone-on-NBN masking pattern (see text) for two different masker levels. The target is a 100-Hz wide NBN centered at 2 kHz. Circles (40 dB masker) and squares (60 dB masker) present data taken from Zwicker (1980, Fig. 2). Model predictions are indicated by lines. Target thresholds are expressed in dB SPL.

The third type of masking pattern considered here involves masking of NBNs by a tone. Zwicker (1980) measured masking of a 100-Hz wide NBN target centered at 2 kHz by tones of constant level and various frequencies. In Fig. 2.9, two of his masking curves are reproduced, obtained at 40 (circles) and 60 dB SPL (squares). No data points were obtained around 2 kHz. Since target level at threshold is plotted against *masker* frequency, and not against *target* frequency (as in Figs. 2.7 and 2.8), the plot shows a mirrored masking pattern. Again, reversion of asymmetry with level is observed and, again, this asymmetry is inadequately described by the model (lines in Fig. 2.9). Moreover, at small target-masker distances, the predicted thresholds are systematically higher than the measured thresholds. This discrepancy also holds for a NBN target centered on a tonal masker, as we know from own unpublished data for comparable masking conditions.

This overestimation of on-frequency masking is probably due to the use of noise targets. The interaction of a tonal masker and a noise target gives rise to fluctuations in the signal envelope (similar to beating effects in tone-on-tone masking), which are likely to be more easily detectable than the overall intensity change. Since the model's detection criterion is based on discrimination of the latter type, it predicts too much masking.

Summarizing the observations on masking patterns, we can state that the excitation-pattern model is too linear when applied to masking with a tonal masker. Since such a discrepancy is not observed for masking patterns of a NBN masker, we can conclude that the nonlinearities in NBN-on-tone masking are less pronounced than they are in tone-on-tone masking. Furthermore, the model overestimates on frequency masking of noise targets by tones.

2.3.2 Growth-of-masking functions

Nonlinear effects in tone-on-tone masking are clearly observed when the masker and target frequencies are held constant, and target threshold is measured at a number of different masker levels. A plot in which the target level at threshold is given as a function of masker level is called a growth-of-masking function.

Schöne (1979) presented data of this type. He obtained thresholds of continuously-presented tonal targets in the presence of a continuous tonal masker by the method of adjustment. He added low-pass noise to mask distortion products. Various target-masker frequency distances (expressed in Bark) were used and the results (expressed in dB SL) for each distance were averaged over different masker frequencies. Predicted growth-of-masking functions are presented in Fig. 2.10 together with Schöne's data. Data for identical masker and target frequencies ($\Delta z = 0$) are included in both panels.

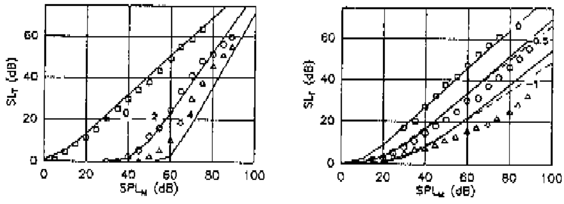

Figure 2.10: Masking functions for tonal maskers and tonal targets measured by Schöne (1979, Figs. 1, 2b). Data are represented by symbols, model predictions by lines. In the left panel the masker frequency is below the target frequency. Masking functions for three different target-masker differences are plotted; $\Delta z = 0$ (squares, upper curve), 2 (circles, middle curve), and 4 Bark (triangles, lower curve). In the right panel the masker frequency is higher than the target frequency. Again, growth-of-masking functions for three different target-masker distances are plotted; $\Delta z = 0$ (squares, upper curves), -0.5 (circles, middle curves), and -1 Bark (triangles, lower curves). Solid lines are calculated with fixed upper filter slopes; dashed lines show results obtained with level-dependent upper slopes (see text).

For masker frequencies f_M below the target frequency f_T, both the model and data show a nonlinear, expansive behavior (circles and triangles in the left panel of Fig. 2.10). Because of the logarithmic measures, nonlinearity shows up as a line with a slope different from unity. The data of Schöne (data points in the left panel of Fig. 2.10) show that the steepness of the growth-of-masking functions at intermediate masker levels (dB target level increase per dB masker level increase) increases with masker-target frequency difference Δz, ranging from 0.95 dB/dB at $\Delta z = 0$ Bark to 2.1 dB/dB at $\Delta z = 4$ Bark. At high masker levels ($>$ 75 dB SPL) the slope of the measured growth-of-masking functions decreases and comes close to unity. The model simulations (solid lines in the left panel of Fig. 2.10) also show nonlinear growth that increases with frequency separation. However, the model underestimates the upward spread of masking, especially at large f_M - f_T differences. Furthermore, the slope of the predicted threshold does not decrease at high target thresholds. For masker frequencies above the target frequency, the data show a strong compressive nonlinearity, increasing with $f_M - f_T$ (data points in the right panel of Fig. 2.10). The model (solid lines in the right panel of Fig. 2.10) shows some compression, which is due to the increasing role of off-frequency listening at higher levels (cf. Fig. 2.5). Compared with the data, the model is far too linear. As mentioned earlier, this suggests that the upper slopes of the auditory filters should become steeper with increasing level. Glasberg and Moore (1990) suggested that the upper slope increased slightly with level at medium center frequencies (as summarized in the review by Moore and Glasberg, 1987b), but the effect of level was less consistent at very low and high center frequencies. For simplicity, they assumed that the upper slope did not vary with level. Using the level dependence proposed by Moore and Glasberg (1987b), we have repeated the calculations simulating Schöne's experiment. The results are shown in the right panel of Fig. 2.10 by dashed lines. Although there is some improvement, the model still predicts far too much masking in a situation with $f_M > f_T$.

masker frequency (Hz)	measured slope of growth-of-masking function	standard deviation	calculated slope of growth-of-masking function
690	1.55	0.26	1.43
800	1.28	0.04	1.27
909	0.93	0.05	1.07
1000	0.89	0.05	0.93
1052	0.67	0.08	0.95
1111	0.59	0.07	0.86
1250	0.26	0.07	0.75

TABLE 2.1. Comparison of slopes of growth-of-masking functions as reported by Bacon and Viemeister (1985, Table 1.) and slopes calculated by the model. The data are averaged over three subjects; the standard deviation is indicated in the third column. The target was a 1-kHz tone with a duration of 20 ms. The maskers were continuous sinusoids with frequencies as indicated in the first column.

The growth-of-masking functions of Bacon and Viemeister (1985), obtained with short (10 ms) 1-kHz targets in continuous tonal maskers of various frequencies, show the same trend as Schöne's data: expansive and compressive nonlinearities for $f_M < f_T$ and $f_M > f_T$, respectively. The measured and modeled slopes of masking are listed in Table 2.1. The model reasonably reproduces the expansive nonlinearity for $f_M < f_T$, but for $f_M > f_T$ it again shows too little compression. The discrepancy is largest for large f_M - f_T differences.

We now consider a growth-of-masking function obtained with a noise masker. Kohlrausch and Langhans (1992) measured masking of a 800-Hz tone by a 100-Hz wide band of noise centered at 450 Hz at various levels. The resulting growth-of-masking function for monaural signal presentation, averaged over three subjects, is reproduced in Fig. 2.11 (squares). As in Schöne's tone-on-tone data, there is an expansive nonlinearity. However, the data show no return to linearity at high signal levels. The same figure shows the model predictions and these are in excellent agreement with the data.

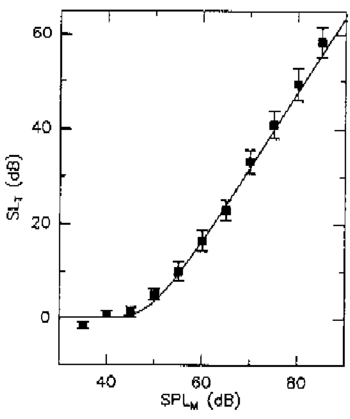

Figure 2.11: NBN-on-tone growth-of-masking function for a target-masker distance of 2.8 Bark. The masker was a 100-Hz wide NBN around 450 Hz; the target was a tone of 800 Hz. Data (solid squares with error bars) are taken from Kohlrausch and Langhans (1992, Fig. 4). The model prediction is indicated by the solid line.

The observations on growth-of-masking functions are consistent with the observations on masking patterns, described above. First, the data show a different nonlinear behavior depending on the type of masker (tonal versus narrow-band noise masker). Second, the model does not show this difference, and its predictions agree much more closely with the data obtained with a NBN masker than those obtained with a tonal masker. It is not surprising that a purely spectral model, based on auditory filters, does not discriminate between tonal and noise maskers of subcritical bandwidth. The difference between the two is simply not resolved. The fact that actual psychoacoustical data do show such a difference, indicates that explaining masking in terms of spectral resolution alone is an oversimplification. Temporal effects that evolve from masker fluctuations could play a role in the observed difference. In this context it is interesting to note that the RoEx filter form of the auditory filters is deduced from masking data with (notched) noise maskers.

2.3.3 Psychoacoustic tuning curves

Unrestricted tuning curves

A psychoacoustic tuning curve (PTC) is obtained by measuring the level of a tone or narrow-band noise needed to just mask a target of fixed level and frequency, as a function of masker frequency. When a low-level target is used (at 10 or 20 dB SPL) the curves show resemblance to tuning curves of auditory nerve fibers, obtained in physiological experiments (e.g., Zwicker, 1974). The curves have a

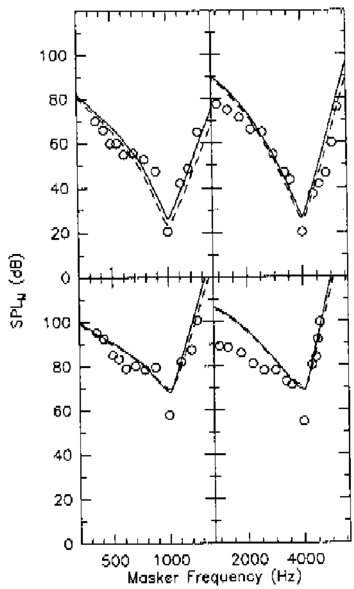

Figure 2.12: Psychoacoustical tuning curves taken from Nelson and Fortune (1991, Fig. 1 and 2) (circles), and model predictions by both the single-band (solid line) and the multiband (dashed lines) version of the model. The target is a tone at a sound pressure level of 20 dB (upper panels) or 60 dB (lower panels). The target frequency was 1 kHz in the two left panels and 4 kHz in the two right panels.

sharp 'tip' at the characteristic (or target-) frequency, and a long low-frequency tail, showing the asymmetry of tuning and masking, respectively. The resemblance between physiological and psychophysical curves is believed to reflect the fact that only a small number of fibers are excited by the low-level target and that spectral masking of this type is determined by peripheral frequency selectivity. The latter is also a basic assumption of an excitation pattern approach.

We simulated experiments of Nelson and Fortune (1991), who measured PTCs with a 1- and a 4-kHz target at levels of 20 and 60 dB SPL. Maskers were pure tones, ranging from $0.4f_T$ to $1.2f_T$, with a duration of 500 ms. The targets had a duration of 250 ms and were temporally centered in the maskers. The high-level PTCs (target at 60 dB SPL) were measured in the presence of a low pass noise to prevent usage of distortion products. Some PTCs from Nelson and Fortune (1991) are reproduced in Fig. 2.12 (circles). The two upper panels show low-level PTCs (target 20 dB SPL), the lower panels high-level PTCs. The left differ from the right panels in the target frequency used (1 and 4 kHz, respectively). Like most PTCs measured under comparable conditions, the data have a typical 'tip-tail' shape: a sharp minimum at the target frequency and a shallow tail at the low-frequency side (Stelmachowicz and Jesteadt, 1984). When plotted on a relative frequency scale (f_M/f_T), the shapes of the 1- and 4-kHz PTCs are very much the same, showing a constant-quality like tuning. For higher levels, the tuning towards lower frequencies is broadened; this is an indication of upward spread of masking.

The sharp tip could have been somewhat influenced by the occurrence of beats, but this is certainly not the only origin of sharp drop of masker level in the vicinity of the target frequency, since this drop is also observed when the influence of beats is excluded. Such conditions include usage of short (50 ms) targets (Vogten 1978) and noise maskers (Nelson and Fortune, 1991).

The solid lines in Fig. 2.12 are the results of the model simulations. The typical tip-tail form in the data (circles) is absent in the simulations. This is to be expected with RoEx-filters since, apart from off-frequency listening, a modeled PTC is an inverted copy of the auditory filter form. The modeled PTCs at 4 kHz show too little upward spread, resulting in too sharp tuning at the low frequencies. This mismatch is most pronounced in the 4-kHz high-level PTC: modeled thresholds are too high by 15 dB and more. This agrees with the observation made in the last section that the model is behaving too linearly with respect to level dependence, particularly when a tonal masker is used.

Multi-channel detection criterion

Because of the large dynamic range in these measurements, an improvement (or at least a change) of model performance might be expected when an integrative detection criterion is used instead of a local one (see section 2.1.5). With help of quantitative, schematic excitation patterns, combined with an integrative detection criterion, Patterson and Moore (1986) arrive at fairly realistic modeled PTCs. In order to check if an integrative detection criterion leads to an improvement, we adopted the method of Florentine and Buus (1981) to combine the information over different channels, and repeated the calculations of the PTCs with this "multi-band paradigm".

Florentine and Buus (1981) divided the auditory range in 24 bands, equally spaced on a Bark scale. Discriminability of two patterns was evaluated in the following way. In each band the differences in excitation, expressed in dB, were calulated. The decision variable was obtained by squaring and adding these differences in excitation. The rationale behind this procedure is the assumption that d' is proportional to log intensity, and that the information of the different channels is treated as being independent.

We modified this procedure by replacing the sum over fixed bands by a sum over all filters (this avoids discontinuities in frequency dependence). Of course, by the very nature of the Glasberg and Moore model, these filters are equally spaced on an ERB scale, not a Bark scale. The free parameter introduced by this procedure is the scaling of the decision variable (cf. the minimum local change of excitation in the standard procedure). It was calibrated to the nonintegrative version of the model (the "single-band" version) by demanding that both procedures yield the same intensity JND for a 1-kHz tone at 50 dB SPL.

The effect of introducing integration over different channels is shown in Fig. 2.12; the dashed lines are modeled PTCs with the integrative criterion described above. The integration of multiple channel information only causes minor changes in the modeled PTCs, the masker levels at threshold being lowered in nearly all cases. The conditions in which target and masker frequencies coincide (1 and 4 kHz in the left and right panels of Fig. 2.12, respectively), correspond to intensity JNDs. For the low-level PTCs the multiband model predicts a lower masker level at threshold than the the single-band model does. For the high level PTCs this situation is reversed. Translated to intensity JNDs this means that the multiband model introduces an extra decrease of intensity JND with level, superimposed on the increase due to the upward spread of excitation. Of course, this extra decrease originates from the increased *range* of excitation with level. All this is in agreement with the observations of Florentine and Buus (1981). Because the multiband model is calibrated to yield the same intensity JND for a 50-dB-SPL tone at 1 kHz, and in view of its increased nonlinearity, it is obvious that it will predict higher intensity JNDs at 20 dB SPL and slightly lower intensity JNDs at 60 dB SPL.

For masker frequencies differing from the target frequency, the multiband model predicts lower masker levels at threshold. This corresponds to decreased audibility. At first sight it might seem contradictory that multiband detection would lead to more masking than single-band detection, but it is a consequence of the calibration (see above): if the multiband model had been adjusted to match low-level intensity JNDs of the single-band model then it would have shown *less* masking at higher levels.[8] The changes are most marked at the higher flanks of the PTCs; this agrees with the observation that OFL is more important when the masker has a higher frequency than the target (see Sect. 1.2.3). For our present purposes the following two observations are most important. First, the changes introduced by the multiband detection criterion are small; they never exceed 5 dB. Second, they do not lead to an improved resemblance of modeled and measured data. These observations bear out our arguments for considering only the single-band model for the majority of simulations (see method section).

Restricted tuning curves

Johnson-Davies and Patterson (1979) measured PTCs in the presence of a second, fixed masker for the purpose of investigating the possible influence of off-frequency listening. The target was a 2-kHz sinusoid of 20 dB SPL, with a duration of 500 ms, gated simultaneously with the masker(s). The maskers were two NBNs, produced by multiplying a sinusoidal carrier and 50-Hz low-pass noise. The fixed masker was set at a level 10 dB below the level needed to mask the target. Its frequency was either 1.8 or 2.2 kHz. Adding the fixed masker on one side of the target frequency generally had a marked effect on the opposite side of the PTC, decreasing the level of the masker needed to mask the target. The authors interpreted these results qualitatively in terms of an excitation-pattern model (cf. also Patterson and Moore, 1986, pp. 158 ff.).

Threshold calculations with the excitation-pattern model are presented in Fig. 2.13 together with the data of Johnson-Davies and Patterson. Open triangles in all panels in Fig. 2.13 represent 'ordinary' PTCs, measured without an additional masker. These data are the same in all three panels. A comparison of these data with the model description (dashed lines) reveals the same discrepancy as observed in Fig. 2.12; the absence of a sharp tip in the modeled PTC which generally leads to an underestimation of the actual frequency selectivity.

In the left panel Fig. 2.13 data and simulations are presented for PTCs with and without an additional, fixed masker below the target frequency. Data are indicated by triangles, model calculations by lines. The PTC measured without the extra masker is indicated with open triangles (data) and dashed lines (simulation). Addition of the fixed masker at 1.8 kHz gives rise to a downward shift of the upper slope of the PTC in both data (filled triangles) and simulations (solid lines). In the model this shift increases with frequency, which can be understood from the increasing role of OFL with masker level (see Sect. 2.2.3). This trend is also seen in the data; adding the fixed masker decreases the upper slope of the PTC. In the

[8] It appears that the most marked differences caused by the introduction of the multi-channel criterion are in fact a result of the calibration. This could have been avoided by using the detection criterion based on intensity JNDs at all levels, as proposed by Fletcher (1930, see Sect. 1.3.2). This would, however, have been incompatible with the model of Glasberg and Moore.

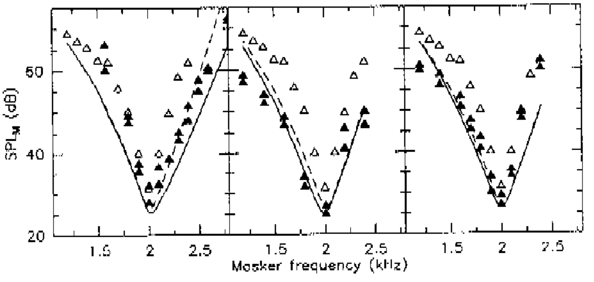

Figure 2.13: Psychophysical tuning curves measured with and without an additional fixed masker. Data points taken from Johnson-Davies and Patterson (1979, Fig. 3, subject DJD). The ordinate is the total power of the variable noise masker. Lines are model predictions. Unrestricted PTCs are indicated with open triangles and dashed lines, respectively (they are identical in all three panels). Restricted PTCs are indicated with closed triangles and solid lines. The target was a tone of 2 kHz and a sound pressure level of 20 dB. Maskers were 100-Hz wide NBNs. The panels differ in the fixed masker that is applied: In the left panel, the fixed masker has a center frequency of 1.8 kHz, and its level is set 10 dB below the level needed to mask the target. In the middle and right panels the fixed masker has a center frequency of 2.2 kHz. Its level is set to respectively 10 and 15 dB below the level needed to mask the target.

model, however, the influence of the extra masker is far less pronounced than in the data, where it exceeds 20 dB in some conditions.

The middle and right panels of Fig. 2.13 illustrate the effect of an additional fixed masker at a frequency above the target frequency. When a fixed masker is added at 2.2 kHz with a level 10 dB below the level needed to mask the target, the data (filled triangles in the middle panel of Fig. 2.13) show a marked effect of this extra masker. Over the whole range of the PTC, masker levels at threshold decrease, the amount of decrease being greater at large frequency separations between target and masker. The effect is that the PTC is shifted to lower levels and broadened. Data are the filled triangles in Fig. 2.13. When the level of the fixed masker is decreased (15 dB below the level needed to mask the target), the effect on the higher side of the PTCs almost disappears (filled triangles in the right panel of Fig. 2.13). In this case, the influence on the lower side is also less pronounced, but the reduction is still 5 dB for a masker at 2.1 kHz.

The model does not predict such a dramatic effect of the extra masker; as expected from the discussion in Sect. 2.2.3, the role of OFL in the model is small when the frequency of the variable masker is below the target frequency (middle and right panels of Fig. 2.13, solid curves). The presence of the fixed masker causes a small, constant shift of a few dB in the lower flank of the PTC. In the data, the influence of the extra high-frequency masker is much greater; shifts of 10-15 dB occur when the fixed masker is added. The model also fails to describe the shift of the higher slope as seen in the middle panel Fig. 2.13, for which the authors do not offer an explanation. Clearly, it cannot be caused by OFL, since this will only affect the opposite side of the PTC.

An explanation for the large discrepancy between the PTC data and the model can be sought in two directions:

- the RoEx filters are not a good description of auditory filters in this conditions. As described in Sect. 2.2.3, the role of off-frequency listening is strongly dependent on the exact shape of the filter.

- the data of Johnson-Davies and Patterson (1979) are not solely determined by off-frequency listening. The shift of the upper slopes in the middle panel of Fig. 2.13 indicates that there is also a nonlinear interaction between the two maskers that is not due to off-frequency listening alone (cf. Humes and Jesteadt, 1989).

2.3.4 Two-tone masking

One of the classical procedures for estimating the critical bandwidth was introduced by Zwicker (1954). He used two equally intense tones to mask a NBN located midway in frequency between the two maskers. The experimental parameter was the spectral separation Δf between the two masker components. No precautions against the use of distortion products are reported, but in view of the moderate masker level it is unlikely that combination tones have significantly affected the measurements. In addition, essentially the same pattern was observed for even lower masker levels of 30 dB SPL per component. For small masker separations, he found a constant threshold value. Above a certain critical separation, he observed a sudden drop in threshold. He identified this critical Δf with the critical bandwidth at the target frequency. A newer set of data is presented in Zwicker (1980), from which the data in Fig. 2.14 are taken (circles). The target was a 100-Hz wide NBN centered at 2 kHz. The masker consisted of two 50-dB tones at frequencies 2 kHz $-\Delta f/2$ and 2 kHz $+\Delta f/2$, respectively. An estimate of the critical bandwidth from the threshold data results in a value of 400 Hz. In the same figure, model simulations are plotted (solid line). The model predicts too high thresholds for small Δf values. This effect has already been described for tone-on-NBN masking in a previous section, where we suggested that envelope fluctuations play a major role in this condition. For

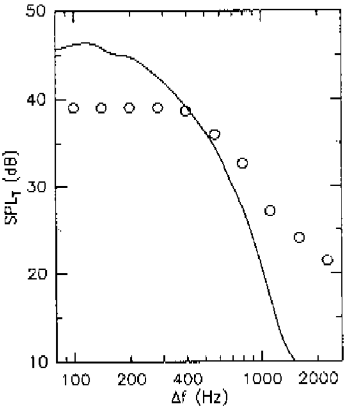

Figure 2.14: Two-tone-on-NBN masking curve as a function of the spectral distance between the two masking tones. The sound pressure level of the masker was 50 dB/tone. The target is a 100-Hz wide NBN centered at 2 kHz. The data (circles) are taken from Zwicker (1980, Fig. 2). The model prediction is indicated by the solid line.

large Δf, the model predicts too low thresholds. This could be related to a lack of upward spread of the model in the case of a tonal masker (see above). Together, these deviations result in a large mismatch between data and model predictions. The occurrence in the model prediction of a gradual decrease of threshold with increasing masker separation resembles the results of Green (1965). This can hardly be taken as an argument in favor of the excitation pattern model; as argued by Green (1965), his data are considerably influenced by temporal effects, for which a model is presented in his paper.

We now consider how well the critical bandwidth can be derived from two-tone masking curves. At first glance one would expect the -3-dB point of the two-tone masking functions, calculated with the excitation pattern model, to be directly related to the -3-dB points of the modeled auditory (RoEx) filters. To verify this, we calculated two-tone masking curves of the type described above, for masker levels of 20, 40, 60, and 80 dB SPL/tone. Fig. 2.15 shows the Δf_{2t} value of the -3-dB points of the masking curves (closed circles) as a function of the masker level. In the same figure we plot the equivalent rectangular bandwidth (ERB) of the Roex filter at 2 kHz (squares), and the -3-dB bandwidth Δf_{-3dB} of the same RoEx filter (open triangles). A comparison between the curves shows that, when applied to the

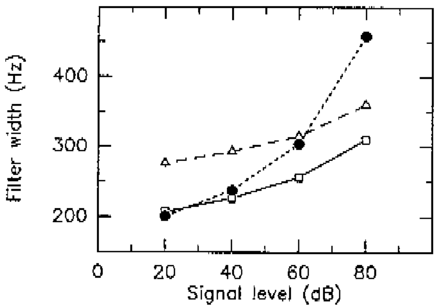

Figure 2.15: Different measures of the bandwidth of the 2-kHz auditory filter, used in the excitation pattern model, plotted as a function of signal level: Equivalent Rectangular Bandwidth of the RoEx filter (squares), -3 dB bandwidth of Roex filter (triangles), and bandwidth as estimated from a simulation of two-tone masking (circles).

model, two-tone masking is an unreliable method for recovering the bandwidths of the auditory filters in the model.

The reason for the excessive growth of Δf_{2t} with signal level is the following. With small masker separations, the target is not resolved by the filters. Detection is then solely based on intensity discrimination, which, in terms of the present model, is performed most effectively at the upper slope of the excitation pattern. Owing to the expansive growth of these upper slopes, this cue becomes increasingly strong with increasing masker level. On increasing the masker separation, the combined excitation pattern of the two tones shows a minimum in the center and detection of the target will occur at the filter centered at the target frequency. Since the filters broaden at higher levels, this second cue becomes weaker with increasing level. The drop in the masking curve results from the competition between these two cues and is not a straightforward reflection of the filter width alone.

It is possible that these complicated interactions between the two cues are just a strange feature of the present model and do not play any role in real masking. On the other hand, one cannot just ignore cues based on upward spread of excitation. This effect has been successfully used to model intensity discrimination (Zwicker, 1956, Florentine and Buus, 1981). In any case, the results shown in Fig. 2.15 warn against too simple an interpretation of experimental data in terms of theoretical concepts.

2.4 Summary and discussion

We tested the applicability of auditory filter shapes to the prediction of masking. For this test, we used the excitation-pattern model of Glasberg and Moore (1990), which is based on RoEx filter shapes. The model gives a fairly good description of masking by narrow bands of noise. When tonal maskers are used, however, the model shows too little asymmetry in masking patterns, and too little nonlinearity. Taking the power spectrum model of masking literally (see Introduction), this would mean that different mechanisms underly the generation of excitation patterns of tones and subcritical noise bands.

Low-level psychophysical tuning curves are poorly described by both the single-band and multiband version of the model. This is probably due to the purely excitatory character of the model, which precludes suppression and active processes in the cochlea (Pickles, 1988; Goldstein, 1990). The effect of an additional, fixed masker when measuring a PTC (Johnson-Davies and Patterson, 1979) is not adequately predicted. Apart from lack of suppression and active cochlear processes, this may also point to incorrect filter shapes in the model because these effects involve OFL, which is dependent on details in the filter shapes (see Sect. 2.2.3).

Finally, two-tone masking is poorly described. This may be due to the fact that the model is too linear. Nonlinearities will dominate the data when more complex signals are used.

In general, it seems that explaining even fairly simple masking situations in terms of just peripheral filtering and T/M ratio is an oversimplification. Both parts of the assumption can be criticized:

peripheral filtering: the actual behavior of the cochlea is not well described by the quasi-linear filter approach, which lacks suppression and other nonlinear behavior. These deficiences could explain the failure to describe PTCs and tonal growth-of-masking functions properly.

T/M ratio as criterion for detection: this assumption states that a simple power detection is the basis for detection of masked signals. Several experimental reports, however, contradict this view: profile analysis (Green, 1988), comodulation masking release (Hall *et al.*, 1984), and nonlinear addition of masking (Humes and Jesteadt, 1989). Even in the simple case of a NBN masking a tone, there is evidence that simple power detection fails to explain the observed phenomena (Kidd *et al.*, 1989; Richards 1992). The systematic difference between masking by subcritical NBNs and masking by tones also indicates that masking is not just governed by filtering and T/M ratio.

These problems warn against a too literal interpretation of 'auditory filters'. Measuring their detailed shape only makes sense if the results are fairly independent of the method used. The model calculations in this study show that modeling spectral masking by a bank of bandpass filters followed by an power detector is an oversimplification that can lead to large mismatches between experimental data and model predictions. From these mismatches, we suspect that the notched-noise method is too 'special' to yield generally applicable filter shapes. In the first place, using noise as a masker seems to have a linearizing effect on masking, compared to tonal maskers (cf. Sect. 2.3.3). In the second place, masker fluctuations are likely to affect masked threshold, a fact that is not appreciated in the derivation of the RoEx filter parameters. In any case, our observations indicate that the accuracy of predicted masked thresholds based on the RoEx filter parameters is limited.

Chapter 3

The role of envelope fluctuations in spectral masking[1]

Unfortunately, we can test our ideas by experiment.
R. P. Feynman in a lecture on quantum mechanics, April 6, 1962.

Abstract

Two experiments are reported in this study. In experiment I the masking effect of 5 different types of narrow-band maskers was compared. The masker was either a tone, a narrow-band Gaussian noise, or a multiplied noise obtained by multiplying a sinusoid with a low-pass Gaussian noise. The 500-ms noise maskers had bandwidths of either 20 or 100 Hz. All maskers were centered at 1.3 kHz. Five-point masking functions were measured using a 2-kHz tonal target with a duration of 400 ms, temporally centered in the masker. Six subjects participated in the experiment. The data showed several trends. First, the tonal maskers produced more masking than the noise maskers. Second, Gaussian noise maskers produced more masking than multiplied noise maskers of the same bandwidth. Finally, 100-Hz wide noise maskers produced more masking than 20-Hz wide maskers of the same type. Differences in masked thresholds between the various masker types increased with masker level and exceeded 25 dB in some conditions. The results are discussed in terms of masker envelope fluctuations.

In experiment II the masker was a band-pass noise at 1.3 kHz with regular zero crossings and with the envelope characteristics of a 100-Hz wide Gaussian noise. Five-point masking functions were measured using a tonal target of 2 kHz. Masked thresholds produced by this hybrid masker were not significantly different from those produced by a 100-Hz Gaussian masker, but differed significantly from those produced by 100-Hz wide multiplied noise. Therefore, differences in masking between Gaussian and multiplied noise are not due to their different fine structure, but to their different envelope statistics.

[1] Modified version of a paper, written together with Armin Kohlrausch, accepted for publication in *J. Acoust. Soc. Am.*

3.1 Introduction

The masking behavior of narrow-band signals is generally believed to reflect the spectral selectivity of the hearing system (e.g., Schafer et al., 1950). When the masker is a pure tone, however, measurements can be plagued by 'false cues' such as beats and combination tones (e.g., Wegel and Lane, 1924). As a remedy, one often uses a narrow-band noise masker instead of a sinusoid (Egan and Hake, 1950). It is tacitly assumed that, apart from the elimination of the forementioned artifacts, this substitution has no effect on the masked thresholds. Under the assumptions of the so called power-spectrum model of masking (Fletcher, 1940; Patterson, 1974), it is indeed true that equally intense maskers of subcritical bandwidth which are centered around the same frequency, will produce equal masking. This model states that the detectability of a signal in presence of a masker is fully determined by the signal-to-masker ratio after a linear filter process. As long as the masker bandwidth does not exceed the width of the auditory filters, it is obvious that the predicted amount of masking only depends on the masker's level and center frequency.

On the other hand, the substitution of a narrow band of noise for a tone introduces fluctuations in both envelope and fine structure. These fluctuations may play a role in the masking produced by these stimuli. In fact, it has been reported by several investigators that, under suitable spectral conditions, masker envelope fluctuations can cause a *release* from masking (Buus, 1985; Mott and Feth, 1986).

The main goal of the present study is to determine what aspects of a narrow band masker, apart from intensity and center frequency, determine the masker's effectiveness. These 'secondary aspects' include bandwidth, regularity of fine structure, and fluctuations of the envelope. Their importance can be assessed from a comparison between the effect of maskers that differ systematically in these aspects.

As a tool for comparing the different masker types, we have chosen growth-of-masking functions (masked threshold as a function of masker level). There are two reasons for this choice. First, in most studies which directly compare different types of narrow-band noise maskers, masked thresholds are measured as a function of some spectral parameter of the masker or the target (Buus, 1985; Mott and Feth, 1986; Moore and Glasberg, 1987a). Thus, few data are available which directly illustrate the level dependence of the differences between the maskers under consideration. Second, a comparison between data from different studies on narrow-band masking suggested that there exists a considerable difference between tone-on-tone and noise-on-tone growth-of-masking functions (van der Heijden and Kohlrausch, 1994).

In Experiment I, we compare the masking effect of a tonal masker and four different noise maskers that differ in bandwidth and noise statistics. The bandwidth was either 20 or 100 Hz. The statistics were determined by the method of noise generation; the noise masker was either a band-passed Gaussian noise or a multiplied noise, obtained by multiplying a sinusoid with a low-pass Gaussian noise. Both types of band-pass noise have been extensively used in masking experiments. There are two aspects in which Gaussian and multiplied noise differ. First, multiplied noise has regular zero crossings while Gaussian noise has irregular zero crossings. Second, the two types of noise have different envelope distributions: The envelope of Gaussian noise has a Rayleigh distribution, while the envelope distribution of multiplied noise is the positive half of a Gaussian distribution.

Experiment II is designed to discriminate between the two aspects in which Gaussian noise and multiplied noise differ. For this purpose we constructed a 'hybrid' masker which has regular zero crossings but a Rayleigh-distributed envelope. A comparison of the masking effectiveness of this hybrid masker with the two other noise maskers will reveal which aspect of the maskers is responsible for masking differences. If masking functions of the hybrid masker coincide with those of multiplied noise, then clearly the difference in masking is due to the regularity of the fine structure. Conversely, if the hybrid masker and Gaussian noise produce equal masking, then it is the envelope distribution that makes the difference.

3.2 Experiment I

3.2.1 Method

Stimuli

Simultaneous growth-of-masking functions (signal level at threshold versus masker level) were obtained for five types of maskers. The maskers were either a sinusoid (T), a band-passed Gaussian noise (G), or a 'multiplied noise' (M), obtained by multiplying a sinusoid with a low-pass Gaussian noise (the bandwidth of this stimulus is two times the bandwidth of the original low-pass noise). The noise maskers G and M had bandwidths of either 20 or 100 Hz and were centered at 1.3 kHz. For each masker type, masked thresholds were measured at masker levels of 60, 66, 72, 78 and 84 dB SPL. All maskers had a duration of 500 ms. The signal was a 2-kHz sinusoid with a duration of 400 ms, temporally centered in the masker. Both the target and the masker were provided with 20-ms Hanning ramps. A band-passed Gaussian noise with a lower cut-off frequency of 500 Hz and a higher cut-off frequency of 800 Hz was added to each masker interval to prevent subjects from using distortion products at 600 Hz (cubic difference tone) and 700 Hz (quadratic difference tone) as a detection cue. Its total level was always 25 dB below the total level of the 1.3-kHz masker. Data of Zwicker (1979) indicate that this level is sufficient to mask the forementioned distortion products. This was confirmed by a series of pilot studies.

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. Before each run a 4-second circular noise buffer was calculated according to the masker specifications of the run. From this buffer, 500-ms samples were drawn randomly from that buffer for each stimulus. A new buffer was calculated for each run.

Band-limited Gaussian noise was produced as follows. First, a 4-second buffer of wide-band Gaussian noise was obtained by drawing independent samples from a Gaussian distribution. A discrete Fourier transform was applied to this buffer (leading to a spectrum with a spacing of 0.25 Hz between the components). The undesired spectral components were set to zero and an inverse Fourier transform yielded a 4-second buffer of band-limited Gaussian noise. In this way, the long term spectrum of the cyclic noise buffer had infinitely steep spectral edges. The multiplication of noise buffers with a sinusoid in order to obtain multiplied noise did not affect the shape of the 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.

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 500 to 6000 Hz as measured at the ear canal entrance.

Procedure

Masked thresholds were determined with a three-interval forced-choice adaptive procedure (Levitt, 1971). Each trial contained three 500-ms observation intervals separated by 200 ms. The masker occurred in all three intervals. The target occurred randomly but with equal probability in one of the three intervals. After the subject's response was collected, a 300-ms pause preceded the next trial. Correct-answer feedback was provided on a computer screen. Each trial block began with the target about 20 dB above masked threshold. After two consecutive correct responses at the same target 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. Thresholds reported in this paper are the averages of three single threshold estimates. If the standard deviation of the first three estimates exceeded 3 dB, more blocks were measured, until the standard deviation of the last three estimates did not exceed 3 dB. These last three estimates were averaged.

Absolute thresholds at 2 kHz were measured using a similar procedure. In this measurement, observation intervals were marked by a very weak 4-kHz tone (about 20 dB above absolute threshold) presented for 200 ms just prior to the intervals.

Subjects were tested in a single-walled sound-attenuating booth placed within a larger sound-attenuated room.

Subjects

Six subjects participated in the experiment. Four subjects A to D were paid an hourly wage for their services. Subjects E and F were the authors. All subjects reported normal hearing. Absolute thresholds at 2 kHz, measured as described above, were 8, 8, 0, 11, 9 and 5 dB SPL for subjects A,B,C,D,E and F, respectively.

3.2.2 Results

Fig. 3.1 presents the averaged growth-of-masking functions (from here: "masking functions") for the different masker types. The error bars in this figure indicate plus and minus one across-subject standard deviation. The inter-subject differences are largest for the conditions with a tonal masker (open circles); the standard deviation for the averaged tone-on-tone masking functions range from 3 to 6 dB (standard deviations of all individual data never exceed 3 dB). The data for noise maskers (squares and triangles) show less inter-subject differences: the standard deviations of these data never exceed 4 dB. Despite the considerable inter-subject differences, the data of all subjects show the same trends. These trends will now be described on the basis of the averaged data.

The masking function for a tonal masker (circles, solid lines) shows a considerably nonlinear, expansive behavior. The steepest part is the middle part of the function: when the masker level is increased from 66 to 78 dB SPL, the threshold increases by 30.3 dB. This corresponds to a slope of 2.5 dB/dB. At the highest and lowest masker levels the function is less steep. These observations agree with those of Schöne (1979), who measured tone-on-tone masking functions for various spectral masker-target separations.

At all levels the tonal masker produces more masking than each of the noise maskers. All four masking functions measured with noise maskers are less steep in the middle range compared to the tonal masking function, and do not show a flattening towards higher masker levels. When the level of a noise masker is increased from 66 to 84 dB, the thresholds increase by 26 to 34 dB, depending on the noise type. This corresponds to a slope of 1.4 to 1.9 dB/dB. At the lowest masker level (60 dB SPL) all types of noise maskers produce the same amount of masking (threshold at 12 dB SPL); the threshold for a tonal masker at 60 dB SPL is somewhat higher (14 dB SPL). The average absolute threshold for the 2-kHz target is 7 dB SPL.

In order to analyze the the differences between the four noise maskers, a four-way analysis of variance was performed on the thresholds for noise maskers. The four factors of the analysis were masker level, noise statistics (Gaussian versus multiplied noise), masker bandwidth, and subject. It was found that all main effects and the interactions of level with each of the other factors were significant ($p < 0.0001$). The root mean square error ('measurement error', as distinct from the 'inter-subject spread' shown in Fig. 3.1) was found to be 2.0 dB. The results of the analysis seem to justify a detailed description of the differences between the various noise maskers.

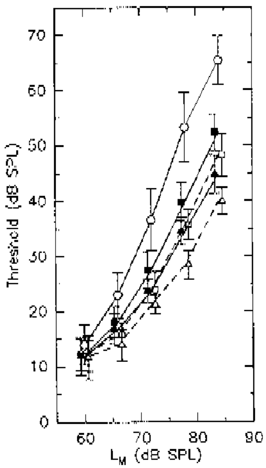

Figure 3.1: Simultaneous growth-of-masking functions measured with various narrow-band maskers centered at 1.3 kHz. The target was a 2-kHz tone. The five curves indicate the results for five different maskers: a sinusoid (open circles), a 100-Hz wide Gaussian noise (filled squares), a 20-Hz wide Gaussian noise (open squares), a 100-Hz wide multiplication noise (filled triangles), and a 20-Hz wide multiplication noise (open triangles). Data points measured with 20-Hz wide maskers are connected with dashed lines. Averaged data of six subjects are plotted; error bars indicate ± one standard deviation across the values of the six subjects. For visual clarity some datapoints have been shifted horizontally; the lines connecting the points, however, have been left unchanged.

At higher masker levels the curves for different noise types diverge and the differences between the various noise types generally increase with masker level. Both bandwidth and statistics of the noise maskers affect thresholds. For both bandwidths, the Gaussian noise produces more masking than the corresponding multiplied noise. For both types of maskers, the 100-Hz wide noise produces more masking than the 20-Hz wide noise. The differences in masking efficiency between tone and noise maskers and between noise maskers among themselves are clearly illustrated by plotting the 'release from masking' for the noise maskers as a function of masker level. This release is defined as the difference in masked thresholds for tonal and for noise maskers of equal level. Fig. 3.2 shows this masking release for the average data. At the lowest masker level all noise maskers show a small release of 2 dB. The amount of release increases with masker level. At the highest masker level (84 dB SPL) the release from masking is 13 to 25 dB, depending on the noise type. The rate of increase is different for the different masker types. At the highest masker levels, the release from masking plateaus because the slope of the tonal masking function decreases, approximating the slope of the noise masking functions.

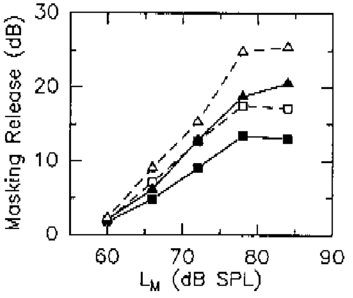

Figure 3.2: Release of masking for noise maskers, plotted as a function of masker level. The ordinate indicates the difference in masked thresholds produced by a tone and a noise masker. The curves are based on the averaged data shown in Fig. 3.1.

3.2.3 Discussion

Before discussing the results of Experiment I, we want to emphasize that the effect of masker fluctuation depends markedly on the spectral relationship between masker and target. In the experiments of the present study the target frequency is well above the masker frequency. When this is not the case, the introduction of masker fluctuations can actually cause an *increase* of masking. This increase has been reported for target frequencies above, but close to the masker frequency (Johnson-Davies, 1981), for targets centered in narrow-band noises (Bos and de Boer, 1966; Hartmann and Pumplin, 1988), and for target frequencies below the masker frequency (Mott and Feth, 1986; Wright, 1992). When the target frequency is close to the masker frequency, interactions between masker and target can play a role in the observed effects.

In the present experiment, the spectral distance between masker and target is sufficiently large (about 700 Hz) to make the influence of direct masker-target interactions (beats) very unlikely. In addition, the 1.3-kHz masker is in the region of the 'shallow tail' of tuning curves measured at 2 kHz (Stelmachowicz and Jesteadt, 1984). This means that the amount of masking is fairly insensitive to the masker frequency. Since all maskers used here have the same center frequency, and the spectra are only different in terms of bandwidth, spectral influences on the observed masked thresholds would in any event be second-order effects. The fact that both spectral effects and direct masker-target interaction play a minor role in the observed masking differences, points to the role of differences in the character of masker fluctuations. Before investigating this aspect in detail, we will now briefly discuss the various properties of the results of Experiment I, and relate them to relevant studies reported in the literature.

The fact that masker fluctuations (both deterministic and stochastic) can give rise to a release from masking, has been reported by several authors (Buus, 1985; Mott and Feth, 1986; Moore and Glasberg, 1987a). Buus (1985) observed the effect with both deterministic maskers (two-tone complexes) and stochastic maskers (narrow-band noise). For both masker types he used targets with a frequency at least 1.5 times the frequency of the highest masker component. He pointed out that the release from masking occurring with a stochastic masker is at odds with energy-detection models of masking (e.g., Green and Swets, 1966), which predict an *increase* of masking due to the introduction of masker uncertainty.

The differences in masking between the various noise maskers mainly appear as different slopes of the masking functions. This level dependency of masking differences is clearly demonstrated in Fig. 3.2, which shows the increase of masking release (taking tone-on-tone masking as a reference) with the level of the different noise maskers. A similar level dependency has been reported by Buus (1985), who compared masking by two-tone complexes and tonal maskers. The tonal masker and the upper component of the two-tone masker had the same frequency of 1075 Hz. For a masker level of 70 dB SPL, the maximum release from masking was 10 dB, whereas with a masker level of 90 dB SPL the maximum release was 25 dB. This across-level comparison, however, is somewhat obscured because he used different target frequencies for the different masker levels: the target frequency was 1600 to 2150 Hz for the 70-dB masker and 3400 to 4800 Hz for the 90-dB masker. The level dependency of the masking release is also consistent with results reported by Moore and Glasberg (1987a). They used 1-kHz sinusoids, amplitude modulated at a rate of 8 Hz, with various modulation depths, to mask a tonal target with a frequency of 1.8 kHz. Three masker levels of 70, 75 and 80 dB SPL were applied. For a modulation depth of 0% (tonal masker) masked thresholds increased by as much as 24 dB when the masker level was increased by 10 dB. For the 100% modulated maskers, masked threshold only increased by 8 dB.

Due to a lack of data in the literature that focus on level dependence, we have not found studies reporting the "saturation" of masking release at high masker levels (above 80 dB SPL) as apparent from Fig. 3.2. This effect, however, might be inferred indirectly by comparing tone-on-tone masking data with narrow-band-noise-on-tone masking data from different studies, as has been done in van der Heijden and Kohlrausch (1994).

As the various growth-of-masking functions obtained in the present study differ in their slope, they cannot be transformed into each other by either a horizontal or a vertical shift. The former would correspond to a constant difference in "effective masker level," the latter to a constant amount of masking release (with a lower limit set by the absolute threshold of the signal). A more detailed discussion of the possible connections between the different growth-of-masking functions can be found in Sect. 3.4.

The effect of bandwidth agrees with Buus' (1985) finding that the release from masking due to beats decreases with increasing beat rate. The highest rate at which Buus (1985) observed a release was 160 Hz. It is also consistent with many data on masking-period patterns, which show a deeper "modulation" of the masked threshold at low repetition rates of a pulsed or modulated masker (e.g., Zwicker, 1976). The bandwidth effect suggests a limited temporal resolution of the detection process. It is improbable that this limitation is due to filtering in the inner ear; an auditory filter centered at 2 kHz with an estimated bandwidth of 240 Hz (Glasberg and Moore, 1990) is unlikely to significantly affect the envelope modulations of a 100-Hz wide masker. A thorough evaluation of temporal effects would include aspects of nonsimultaneous masking, duration effects and modulation detection (cf. Fastl, 1975; Zwicker, 1976) and is beyond the scope of this paper which concentrates on the statistics of masker envelope fluctuations.

A difference in masking efficiency between multiplied noise (M) and Gaussian noise (G) has been reported previously by Mott and Feth (1986). They compared masking patterns of a tone, narrow-band noise with a flat temporal envelope (FM noise), Gaussian noise and multiplied noise. All maskers had a level of 75 dB SPL and were centered around 1500 Hz; they had bandwidths between 50 and 60 Hz. These four maskers have the following characteristics of fine structure and envelope. For the tone and the FM noise, the envelope is constant, while for the other two noise types, it is stochastically varying. The fine structure (i.e., the temporal distance between subsequent zero crossings) is regular for the tone and the multiplied noise, while it is fluctuating for the FM and the Gaussian noise. Mott and Feth (1986) found no significant difference in masking between the tone and the FM noise. In the region from 2 to 3 kHz the thresholds were highest for the tonal masker, about 10 dB lower for the Gaussian noise, and another 5 dB lower for the multiplied noise. These observations agree qualitatively with those of the present study.

Mott and Feth (1986) attributed the difference in masking between M and G noise to differences in fine structure; the multiplied noise has regular zero crossings while the Gaussian noise does not. By concentrating on this aspect, they ignored another distinction between the two types of noise, namely their envelope statistics. The envelope of Gaussian noise has a Rayleigh distribution, whereas the envelope distribution of multiplied noise is the positive half of a Gaussian distribution with zero mean (the envelope being the absolute value of the low-passed Gaussian noise used as the multiplicator). The latter has a maximum at zero (corresponding to the zero crossings of the low-passed noise used as the multiplicator), while in Gaussian noise the probability that the envelope is zero vanishes (cf. Fig. 3.3). Any strategy

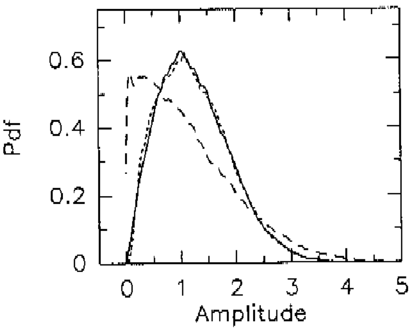

Figure 3.3: Probability distributions of the envelope of three types of noise: Gaussian noise (solid line), multiplication noise (long dashes), and 'hybrid noise' (see text, short dashes). The curves were calculated by averaging the envelope distributions of 100 independent 1-second samples of noise.

using minima in the envelope to improve detection will be sensitive to the differences of envelope statistics. This will be true regardless of the exact nature of the proposed strategy. In the second experiment reported in this paper, we investigate which aspect of the noise makes M a less effective masker than G: its different envelope distribution or its regular fine structure.

3.3 Experiment II

3.3.1 Method

Stimuli

Simultaneous growth-of-masking functions (signal level at threshold versus masker level) were obtained for a narrow-band masker (GM) with envelope characteristics of Gaussian band-pass noise, but with regular zero crossings. Apart from the masker type, the present experiment replicated the measurements of Experiment I (including the band-passed Gaussian noise used to mask distortion products) as described in Sect. 3.2.1.

The masker (GM) was calculated as follows: first a 100-Hz wide Gaussian noise centered at 1.3 kHz was calculated. Next the Hilbert envelope of this noise was extracted. Since the spectrum of this envelope is not band-limited (Lawson and Uhlenbeck, 1950, Sect. 3.8), all spectral components of the envelope above 100 Hz were set to zero. Finally the low-passed envelope was multiplied by a 1.3-kHz sinusoid.

The envelope distributions of Gaussian noise, multiplied noise and 'hybrid' noise are plotted together in Fig. 3.3. It can be observed from this figure that the envelope distribution of the GM masker is practically identical to that of Gaussian noise. In addition, the envelope *spectra* of G and GM noise are practically identical.

This identity of envelope characteristics can only be achieved at the cost of differences in the long-term power spectra. The long-term spectrum of the GM masker has the following properties. It is symmetric around the center frequency of 1300 Hz and has a total width of 200 Hz. The spectrum has a peak at the center frequency and decays approximately linearly on a power scale (cf. Lawson and Uhlenbeck, 1950, Fig. 3.14). It is unlikely that these spectral differences as such influence the masking behavior of the noise. First, the differences are confined to a spectral range between 1200 and 1400 Hz, whereas the signal frequency is 2 kHz. Second, as argued in Sect. 3.2.3, symmetrical spectral changes in the masker would in any case be second order effects in the current setup of the experiment[2].

Procedure and subjects

Methods of data collection and processing are identical to those of Experiment I. Three of the subjects that participated of Experiment I (subjects D, E, and F) participated in this experiment.

3.3.2 Results and discussion

Fig. 3.4 presents average masking functions for GM maskers, combined with data taken from Experiment I. These reference data are masking functions of subjects D, E, and F, measured with 100-Hz wide G and M maskers (cf. Fig. 3.1). For all three subjects the GM masking function coincides with the G masking functions more than it does with the M masking function. A three-way analysis of variance was performed on the results of Experiment II together with the 100-Hz bandwidth data

[2]The choice of cutting off the envelope spectrum above 100 Hz is a compromise between affecting the waveform spectrum of the noise and affecting its envelope distribution. Generally, it is not possible to generate a stimulus that unifies a predescribed spectrum with arbitrary envelope characteristics. The problems encountered here are analogous to the problem of generating spectrally flat noise with a flat temporal envelope ("low-noise noise," Hartmann and Pumplin, 1988).

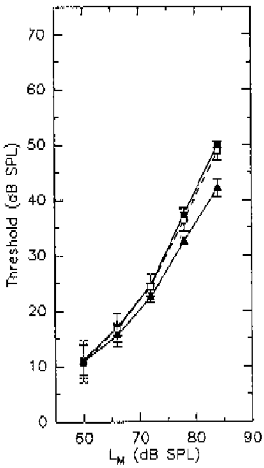

Figure 3.4: Simultaneous growth-of-masking functions measured with three narrow-band maskers centered at 1.3 kHz. The target was a 2-kHz tone. The three curves indicate the results for three different maskers: 100-Hz wide Gaussian noise (filled squares), 'hybrid noise' (see text, open squares), and 100-Hz wide multiplication noise (filled triangles). Averaged data of three subjects are plotted; error bars indicate ± one standard deviation.

of Experiment I. The three factors of the analysis were masker level, envelope and fine structure (the hybrid masker GM fell in the same envelope class as the Gaussian noise but in the same fine structure class as the multiplied noise; Gaussian and multiplied noise differed both in envelope and fine structure). The analysis showed a significant effect of envelope ($F(1,120) = 32.79, p < 0.0001$) and its interaction with level ($F(4,120) = 8.24, p < 0.0001$). Fine structure and its interaction with level were not significant ($p > 0.5$). These observations clearly indicate that differences in envelope statistics rather than differences in fine structure are responsible for the differences in masking produced by Gaussian and multiplied noise.

3.4 Analysis and general discussion

In Experiment II it is found that differences in envelope statistics rather than fine structure are responsible for the different masking behavior of Gaussian and multiplied noise. This being established, we can now summarize the factors that determine the masking differences between the various narrow-band maskers used in the present study:

- level: differences between the various masker types increase with masker level; at low masker levels the masked thresholds converge.

- bandwidth: masked thresholds increase with masker bandwidth.

- envelope statistics: the occurrence of marked envelope minima makes a masker less effective.

Although the fluctuating maskers of the present study are stochastic, it should be noted that each of the listed effects has its equivalent in conditions with deterministic maskers. As for the the level and bandwidth effects, such equivalences have been discussed in Sect. 3.2.3 The influence of envelope statistics is analogous to the effect of modulation depth on the amount of masking by an amplitude-modulated masker (Moore and Glasberg, 1987a). These equivalences indicate that an explanation of the effects should not be dependent on the *stochastic* nature of the fluctuations.

As an illustration of the influence of masker envelope statistics, let us consider the following over-simplified method for predicting masking by a fluctuating masker from masking by a tone. It is very similar to the procedure applied by Fastl (1975), but unlike his approach we ignore forward and backward masking effects. First, the time-varying envelope of the masker is extracted. At every instant the masker is assumed to produce an 'instantaneous masked threshold.' This threshold is simply the masked threshold for a tonal masker at a level equal to the instantaneous power of the fluctuating masker. These masked thresholds are derived from the averaged tone-on-tone masking data in Fig. 3.1 by a simple linear interpolation. Finally, the averaging of 'instantaneous thresholds' leads to the predicted masked threshold. The rationale behind this procedure lies in the assumption that the masked thresholds are a measure of masker excitation and that the threshold is determined by the average excitation of the fluctuating masker (Zwicker, 1970). Mathematically, the procedure boils down to first transforming the instantaneous intensity of a fluctuating stimulus before averaging it. The scale in which the 'instantaneous masked thresholds' are expressed and averaged is crucial for the predictions; like Fastl (1975) we express them in decibels. Compared with a power scale this attaches more weight to the envelope minima.

The procedure just described is an oversimplification in the sense that it focuses on the envelope statistics and ignores effects of temporal resolution. The procedure predicts the same thresholds for two maskers independent of their bandwidth and *rate* of envelope fluctuations, as long as their envelope *statistics* coincide. The results should, therefore, be compared with data obtained with small masker bandwidths, where limitations in temporal resolution should have the smallest effect on detection. In Fastl (1975) the model predictions agreed reasonably well with data for a 32-Hz wide masker (for this bandwidth, the masker fluctuations are too slow to be smeared out by forward and backward masking). In any event, the just proposed procedure should not be regarded as a general model to predict masking by fluctuating maskers; we only introduce it here in order to demonstrate the effect of nonlinear scaling when applied to fluctuating stimuli.

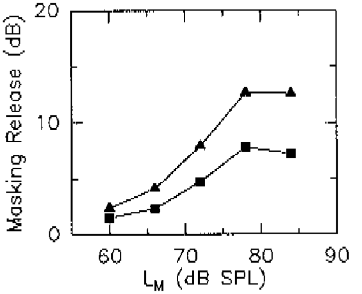

Figure 3.5: Predicted release of masking for noise maskers, plotted as a function of masker level. The ordinate indicates the difference in masked thresholds produced by a tone and a noise masker (cf. Fig. 3.2). The curves show the prediction of the transform-and-average model described in the text. Triangles indicate the release for multiplication noise and squares the release for Gaussian noise.

Results of the prediction are shown in Fig. 3.5, where the difference between calculated thresholds for noise maskers and tonal maskers (masking release) is plotted (cf. Fig. 3.2). The model predictions agree with the data of Fig. 3.2 in that a release from masking by the fluctuations is predicted, which is greater for multiplied noise than for Gaussian noise, and which shows a "saturation" at masker levels above 80 dB. But the predicted amount of release is much smaller than in the data of Fig. 3.2, both for the 20 and the 100-Hz bandwidth data: for Gaussian noise the predicted maximum release is 8 dB and for multiplied noise it is 13 dB. The experiments for a 20-Hz bandwidth show releases of 18 and 25 dB, respectively. The 100-Hz bandwidth data show releases of 13 and 21 dB, respectively. Despite this quantitative disagreement, the calculations demonstrate that a simple transform-and-average model is able to describe a release from masking due to envelope fluctuations. If the transforming part is compressive (i.e., if it disproportionally favors the envelope minima), the release will be greatest for maskers such as multiplied noise, which show distinct envelope minima. Apparently the amplitude-to-dB transformation followed by the masker-to-target-level substitution described above, does not appreciate the envelope minima sufficiently. It is an exercise in curve fitting to find an even stronger compressive transformation that would 'repredict' the 20-Hz data of Fig. 3.2.

In the light of the results shown in Fig. 3.5, Buus' (1985) remark that "it seems clear that no model which bases detection solely on the statistics of energy (or some transformation thereof) within a single channel can explain [his results]" is slightly misleading. If energy refers to the total energy, it is of course correct. But transforming the instantaneous power *before* averaging it, will generally lead to different predictions for different masker statistics. The average power of a fluctuating masker is just one choice of conveniently describing its strength; it is by no means clear that the average power of a stimulus is a sensible measure of its masking effectiveness. In fact, we feel that the very effect of 'masking release caused by envelope fluctuations' suggests that the masker's average power is not a good measure, and that the effect by itself does not justify Buus' (1985) conclusion that "...the auditory system applies some rather sophisticated multi-channel process to optimize detection".

Since masker envelope fluctuations can play an important role in the amount of masking produced by narrow-band maskers, this aspect cannot be ignored in 'addition-of-masking' studies, where masking by a combination of different narrow-band maskers is measured. Combining two 50-Hz wide noise maskers, Lutfi (1983a) observed an 'excess masking' of 10-17 dB over the amount of masking which would be expected from a simple power addition. In a reaction to this study, Moore (1985) argued that the excess masking should partly be ascribed to the influence of masker fluctuations. It is interesting to note that in the original study by Lutfi (1983a) the maskers were Gaussian bands of noise, whereas Moore (1985) in his replication used multiplied noise, thereby ignoring possible differences in masking effect by these two types of noises. Moore's (1985) statement that 'when two independent narrow-band noises are added together, the minima in the combined waveform are less pronounced than for either alone', is correct for multiplied noise as used in his study, but incorrect for Gaussian noise as applied in the original study by Lutfi (1983a). This is so, because any sum of Gaussian processes is again a Gaussian process, and because any band-limited Gaussian signal has a Rayleigh-distributed envelope. Thus a combination of Gaussian noise bands has the same envelope distribution as either component. It is of course true that addition of spectrally distant bands will introduce *faster* envelope fluctuations, but this is a bandwidth effect and not a matter of envelope distribution.

Another example that illustrates the role of masker statistics of combined maskers can be based on Experiment I of the present study. According to the averaged data presented in Fig. 3.1, both a 100-Hz wide Gaussian noise G of 78 dB SPL and a 100-Hz wide multiplied noise M of 82 dB SPL produce the same masked threshold of 40 dB SPL. Addition of two independent samples of G results in the same kind of Gaussian noise, with a level of 78 + 3 = 81 dB SPL. From Fig. 3.1 we read that this 'G+G' masker produces a threshold of 46 dB SPL. On the other hand, incoherent addition of two independent samples of M (with a 90 degree carrier phase difference) results in a Gaussian noise; this is the so called 'quadrature' method of producing

Gaussian noise. The level of this Gaussian noise is 82 + 3 = 85 dB SPL, and from Fig. 3.1 we find a masked threshold of 54 dB SPL. The difference of 8 dB is entirely due to the *type* of noise of the masker constituents. Models in which the masking effectiveness of a complex is predicted from the masking produced by the individual components, (e.g., Humes and Jesteadt, 1989; Humes and Lee, 1992), will predict the same thresholds for the 'G+G' and the 'M+M' masker.

In summary, our data show that, particularly with high masker levels, attention should be paid to the exact nature of narrow-band stimuli used to mask a target at a frequency above the masker frequency. Statistics and rate of the masker fluctuations have a considerable effect on masked thresholds. These differences are by no means restricted to extremely slowly fluctuating maskers. On the contrary, masker fluctuations are relevant under many common experimental conditions, such as the 'low-frequency tail' of psychoacoustic tuning curves. Unfortunately, in many published papers the influence of the type of noise used as masker (particularly multiplied noise versus Gaussian noise) has generally been neglected.

Chapter 4

The role of distortion products in masking by single bands of noise[1]

We must take into account the possibility of interactions among components, and there is no way to learn about the interactions except through experiment.
J. C. R. Licklider in 1951

Abstract

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 a bandpass noise masker ranging from 1 to 2 kHz. In all the 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 degrees. The phase pattern shifted with increasing masker level. The individual shift for the masker range of 45 dB was between 100 and 200 degrees. 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 generalizablity 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 degrees. An analysis of variance showed that signal phase played a significant role for the high-frequency, but not for the low-frequency masker.

[1] Modified version of a paper, written together with Armin Kohlrausch, submitted for publication in *J. Acoust. Soc. Am.*

4.1 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, 1961a), 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 leads to a second difference between noise-on-tone and tone-on-tone masking.

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, 1995a). 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 towards 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.

4.1.1 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 combi nation 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."

It is assumed that the distortion of an arbitrary collection of primaries can be described by a simple *summation* of distortion products that would be produced by single pairs of components when presented in isolation. The validity of this assumption 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 (1983b) observed that the threshold of a signal at the frequency of the cubic combination band ($2f_l - f_h$) produced by two-narrow band primaries actually decreased when the spectral gap between the two primaries was filled with additional noise.[2] The construction of the aggregate according to Greenwood (1971, p. 532 ff.) 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).

These observations do not present an argument against the *existence* of the combinational aggregate of a single noise band, but they question whether any *quantitative* statements on these distortion products can be inferred from knowledge of distortion products evoked by pairs of narrow-band stimuli. Insight into the more general situation has to come from experimental investigations.

Although the data of both Greenwood (1971) and Zwicker and Bubel (1977) do not seem to disagree with the hypothesis of masking by combination products, they do not present conclusive evidence for the effect either. 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 lacking in masking patterns

[2]The interpretation of Lutfi's (1983b) 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 co-modulated, which could affect the generation of the combination bands (see Sect. 4.2.3).

for single noise bands. The combinational aggregate produced by a single masker band can safely be expected to show a high-pass character below the masker's low cut-off frequency. It 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 a new method for evaluating this influence in a more direct way.

4.1.2 Phase-dependent signal-masker interactions

The approach of the present study is based on the fact that direct signal-masker interactions can affect the audibility of a tonal signal in the presence of a broadband noise. 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 (Pfafflin and Mathews, 1966). 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, 1992b; 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.[3] 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

[3]An exception is the combination of sinusoidal stimuli with harmonically related frequencies. Here, a relative phase can be defined on the basis of a common fundamental frequency, and, indeed, phase effects have been observed in some cases (e.g. Clack *et al.*, 1972).

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.

4.2 Experiment 1: phase effect for a single frozen-noise sample

4.2.1 Method

Stimuli

A single 500-ms sample of Gaussian noise was used as a masker throughout this experiment. It was band-limited 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. Band-limited Gaussian noise was produced as follows. First, a 500-ms buffer of wide-band 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 band-limited 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-degree 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.

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 to D, participated in the experiment. Subjects C and D were the authors. All the subjects had extensive experience in listening tasks.

Pilot study: temporal placement of the signal within the masker

In the case of a broad-band noise masker the size of the monaural phase effect can vary considerably between different temporal placements of a short signal in a frozen sample of noise (e.g. Langhans and Kohlrausch, 1992b). Therefore, we preceded the actual measurements by a selection procedure, in which different temporal placements of the signal were realized by a cyclic shift of the *masker*: the noise stimulus was shifted (rotated) within a cyclic buffer. In this way, the stimulus configuration of the experiment is left unaffected. From previous experiences with this procedure we learned that phase effects, when present, tend to *lower* the masked thresholds. For temporal placements for which no phase effect was found, the threshold was nearly equal to the highest threshold of the 'successful' placements. Thresholds in both 'worst cases' came close to thresholds for running-noise conditions. These observations suggest that the placements yielding the lowest thresholds are most likely to show a phase effect. This idea was exploited in a pilot study in which all four subjects of the main experiment took part.

Masked thresholds of a 900-Hz signal were measured as described above for ten different placements of the signal within the noise buffer. These placements were realized by successive 50-ms cyclic shifts of the 500-ms noise buffer [4]. The masker level was 50 dB SPL. For each condition the threshold was measured twice.

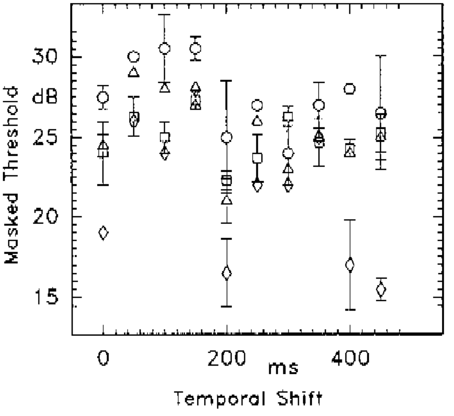

Figure 4.1: The results of the pilot experiment, in which thresholds of a 50-ms tonal target at 900 Hz were measured, masked by a 500-ms frozen sample of bandpassed noise ranging from 1-2 kHz. The error bars show +/- one within-subject standard deviation. The cyclic masker noise buffer was shifted in 50-ms steps as indicated along the abcissa. Different symbols indicate data of different subjects: subject A (circles), B (squares), C (triangles), and D (diamonds). Each symbol presents the average of two data points.

The results are shown in Fig. 4.1. The thresholds varied over a range of 16 dB. Although there were considerable inter-subject 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), 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.

[4]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.

4.2.2 Results

Fig. 4.2 shows the thresholds as a function of signal starting phase for subjects A to D (panels a to d). In each panel thresholds for four different masker levels are shown; different symbols have been used for each masker level. Lines connect

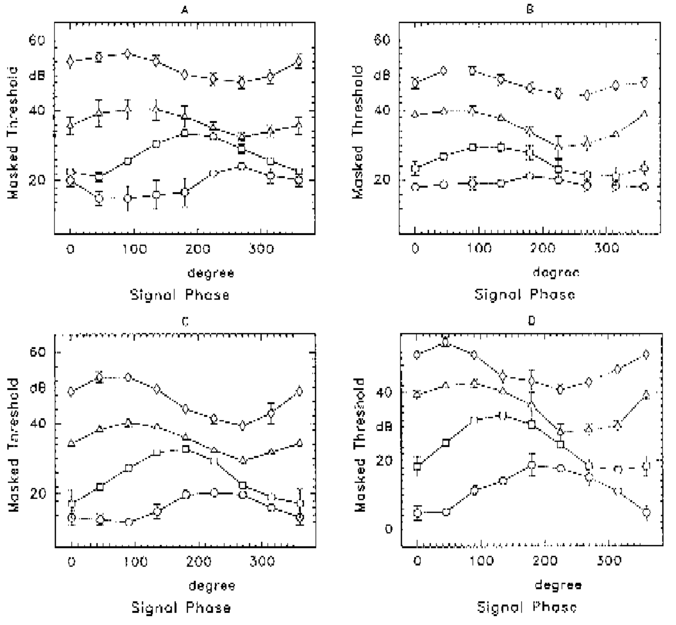

Figure 4.2: Masked thresholds of a 50-ms tone at 900 Hz as a function of signal starting phase. The masker was a 500-ms frozen-noise sample, band-passed from 1-2 kHz. Panels A to D show the results obtained from subjects A to D, respectively. In each panel, thresholds for four different masker levels 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 3 data points; error bars indicate ± one standard deviation.

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 the 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 dB (subject B, 35-dB masker) and 16 dB (subject D, 50-dB masker). The phase effect can be observed for all masker levels (35 to 80 dB SPL). There is an effect of masker level on the phase effect: apart from an expected over-all 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. 4.3 to 4.5.

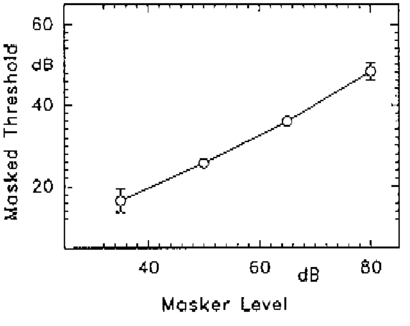

Figure 4.3: Average growth-of-masking function (threshold vs masker level). The thresholds were obtained from the data of Fig. 4.2 by averaging the data for each masker level over all signal phase values and over all subjects. The error bars indicate ± one inter-subject standard deviation.

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. 4.2 over all the phase conditions. In this way, a curve relating average threshold to masker level was obtained for each subject. The curve in Fig. 4.3 is the average of the four individual growth-of-masking functions; the error bars indicate the inter-subject 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 eight points of each of the patterns of Fig. 4.2). This resulted in a curve relating 'spread caused by phase' to masker level for each subject. The average of these four curves is shown in Fig. 4.4; the error bars indicate the inter-subject standard deviation. The phase effect was found to be maximal for a 50-dB masker level. A two-way analysis of

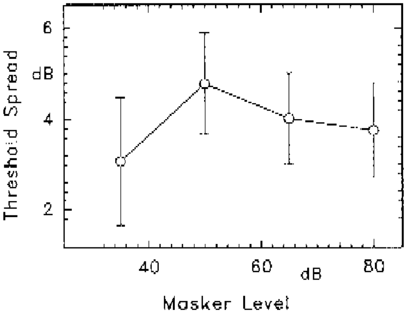

Figure 4.4: The amount of phase-induced variation of thresholds, expressed by the standard deviation of the points in each pattern in Fig. 4.2, plotted as a function of masker level. Each data point presents the value obtained by averaging over all four subjects. The error bars indicate ± one inter-subject standard deviation.

variance (masker level versus subject[5]), 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. 4.2 with increasing masker level we calculated cross-correlation functions between the individual patterns of Fig. 4.2. In these calculations, the curves were treated as being cyclic (with a cycle of 360 degrees). 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 the subjects, now using the 50-dB pattern of subject D as a reference.[6] The offsets that resulted from this procedure

[5]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).

[6]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. 4.2 would be the maximization of the 16th-order cross-correlation function of all the patterns. This was beyond our numerical scope. Since all the 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.

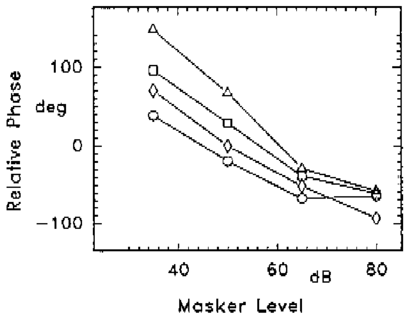

Figure 4.5: Phase of the individual patterns of Fig. 4.2 relative to the 50-dB pattern of subject D, plotted as a function of masker level (see text). Different symbols have been used for the four subjects: subject A (circles), B (squares), C (triangles), and D (diamonds).

were added to the corresponding curves from the within-subject analysis.

The resulting curves are plotted in Fig. 4.5. They show the phase of the individual patterns of Fig. 4.2 relative to the 50-dB pattern of subject D. It can be inferred from Fig. 4.5 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 degrees for subjects A, B, C, and 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, B, C, and D, respectively.

4.2.3 Discussion

Masking by the combinational aggregate

Both the shape and the size of the patterns of Fig. 4.2 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 broad-band 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-degree steps. Most of the threshold-versus-phase patterns reported by Hanna and Robinson (1985,

Fig. 3) have the same sinusoidal shape as the patterns of our Fig. 4.2. Threshold variation caused by signal phase ranged from a few dB to 12 dB. Langhans and Kohlrausch (1992b) 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 Fig. 4.2.

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, 1992b). This involves a straightforward evaluation of the energy or the envelope of that portion of the stimulus that is close to the signal frequency. Langhans and Kohlrausch (1992b) 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[7].

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 flat-envelope part of the signal. 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. 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.

[7]The assumption that the temporary envelope decrease of a fluctuating stimulus is not detectable is consistent with Erdreich's (1977, Sect. 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.

Corrections had to be made for the differences in signal duration and frequency[8]. 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) and Zwicker and Bubel (1977). As 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 over-all 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 over-all 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[9]. 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

---

[8] 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 one-octave difference in signal frequency, corresponding to an approximate doubling of the critical band.

[9] 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.

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 inapplicable to the wider noise bands used in the present experiment, in which the distinction between lower and higher primaries is lost.

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 to 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. 4.3). An explanation for this discrepancy can be sought in two directions. First, level dependencies derived from masking behavior might differ from those measured using the cancellation method as in Zwicker (1979). Support for this explanation can be found in Greenwood (1972), who reported a slightly decreasing CT-to-primary ratio with an increasing primary level. Second, the aggregate can be expected to have a very steep high-pass character. This makes it less self-evident that the thresholds grow linearly with its level, as is the case with spectrally flat masking noise (Hawkins and Stevens, 1950). 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. 4.2, and in analyzed form in Fig. 4.5, 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 degrees 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 degrees 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 degrees 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 to 150 degrees for a 25-dB primary-level increase, depending on the subject.

A detailed comparison 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 primaries. 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. 4.5). Second, individual phase shifts, though conforming to the same trends, show a large variation across subjects (up to 100 degrees over a total range of only a few hundred degrees). At the same time, the levels of the distortion product show little inter-subject 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 inter-subject differences in masked thresholds shown in Fig. 4.2 are more 'horizontal' than 'vertical,' *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 Section 4.2.2 (Figs. 4.3 and 4.5).

4.3 Experiment 2: phase effect for a large number of frozen-noise samples

4.3.1 Rationale

All the 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. The larger number of noise samples necessarily limits the manipulation of the remaining stimulus parameters, thus preventing an elaborate analysis as in Experiment 1. In particular, since we chose to use different sets of noise samples for each of the subjects, a detailed evaluation of inter-subject differences is impossible.[10] On the other hand, the data of many different noise samples allow an objective evaluation of the generalizability of the effect. This is most useful in determining conditions under which the phase effect does *not* occur. In order to test the falsifying capability of this method, a condition was included in this experiment with a noise masker consisting of components *below* the signal frequency. In this situation it is unlikely that masking is determined by combination products (Goldstein, 1967; Greenwood, 1971; Erdreich, 1977).

4.3.2 Method

The procedure was the same as to that used in Experiment 1 except for two modifications. First, the masker duration was decreased from 500 ms 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 Section 4.2.1. 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.

A different set of 20 independent noise samples was used for each subject. For each noise sample, two signal conditions were examined, differing in starting phase by 180 degrees. 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, 2 signal phases, and 2 repetitions). The order of presentation was completely randomized.

The four subjects of Experiment 1 participated in this experiment.

4.3.3 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), performed for the data obtained with the high-frequency (1-2 kHz) masker, 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

---

[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.

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. 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 (Sect. 4.2.1) 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 absence of phase effects 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 Sect. 4.2.3 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 Sect. 4.2.3.

4.4 Summary and general discussion

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 non-overlapping 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 non-overlapping 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 cut-off 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 degrees. Two masker types were used, a high-frequency and a low-frequency masker. A statistical analysis indicated that the effect of signal phase occurred in the case of the high-frequency masker only.

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 towards 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.

Chapter 5

General conclusions

Numerical simulations, combined with a literature study (Chapter 2), led to the following conclusions. The assumption of the power-spectrum model of masking are not consistent with psychoacoustic data. In particular, the estimates of level-dependent auditory frequency selectivity derived from notched-noise experiments (Glasberg and Moore, 1990) are in conflict with the following characteristics of masking:

- the large differences in masking effectivity between tones and noise bands
- the amount of upward spread for tone maskers
- the very shallow growth of masking produced by tonal maskers towards lower frequencies

Furthermore, an analysis of the auditory filters of the model revealed that, in the case of a single narrow-band masker,

- the predicted effect of off-frequency listening on the amount of masking is marginal
- a multi-band detection criterion does not lead to improved prediction of masking.

Experimental work (Chapters 3 and 4) led to the following conclusions with regard to the different masking behavior of noise and tone maskers. The relatively linear masking behavior of noise is a deviation from the 'nonlinear standard' set by tonal maskers, in the sense that it is caused by specific properties of the noise. These specific determinants of noise masking are entirely different for 'upward' and 'downward' masking, *i.e.* masking towards frequencies above and below the masker frequency, respectively.

In case of upward masking, the envelope fluctuations of noise maskers result in a release from masking compared to pure-tone masking. This release from masking increases with masker level, thus yielding a more linear growth of masking than the very steep growth for tonal masking. Both the statistical distribution of the envelope and the rate of its fluctuations affect the amount of release, a maximum release occurring when the envelope has marked minima and when the rate of its fluctuations is low. Fluctuations in fine structure of a noise band have no effect on its effectiveness as a masker. A simple transform-and-average scheme is sufficient to understand masking release caused by masker envelope fluctuations; there is no need for across-frequency mechanisms to explain the effect.

In the case of downward masking by a noise band with a steep lower spectral edge, evidence was presented for the existence and masking potency of distortion products generated by the mutual interaction of noise components. Auditory nonlinearities generate noise components below the spectrum of the original noise. This secondary noise, or 'combinational aggregate' is responsible for the downward masking over a large dynamic range of the masker. Level-induced phase effects in frozen-noise masking showed that the combinational aggregate has characteristics comparable to those of odd-order combination products such as the cubic difference tone. The growth of the aggregate with noise level results in a steeper and thus more linear growth of masking than the very shallow growth observed with a tonal masker.

Since the predictions of the model of Glasberg and Moore are generally more in line with data obtained with noise masker data than those obtained with tonal maskers, it is probable that at least one of the two 'special effects' of noise maskers described above played a role in the notched-noise masking experiments that underly their derived auditory filters. Indirectly, these observations narrow down the range of applicability of the power-spectrum model of masking. More generally, they undermine the degree of reality that can be attached to auditory filters, for if they cannot be reliably measured then they do not exist.

References

Bacon, S. P., and Viemeister, N. F. (**1985**). "Simultaneous masking by gated and continuous sinusoidal maskers," J. Acoust. Soc. Am. **78**, 1220–1230.

Beerends, J. G., and Stemerdink, J. A. (**1992**). "A perceptual audio quality measure based on a psychoacoustic sound representation," J. Audio. Eng. Soc. **40**, 963–978.

Bilger, R. (**1979**). "A revised critical band hypothesis," in *Hearing and Davis: Essays honoring Hallowell Davis*, edited by S. K. Hirsch, D. H. Eldridge, I. J. Hirsch, and S. R. Silverman (Washington University Press, St. Louis, Missouri), pp. 191–198.

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

Bos, C. E., and de Boer, E. (**1966**). "Masking and discrimination," J. Acoust. Soc. Am. **39**, 708–715.

Brandenburg, K., and Sporer, T. (**1992**). ""NMR" and "Masking Flag": Evaluation of Quality Using Perceptual Criteria," in *Proc. of the 11th Int. AES Conf., Portland, Oregon, USA, May, 1992*, edited by D. R. von Recklinghausen (AES, Inc., New York), pp. 169–179.

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.

Egan, J. P., and Klumpp, R. G. (**1951**). "The error due to masking in the measurement of aural harmonics by the method of best beats," J. Acoust. Soc. Am. **23**, 275–286.

Ehmer, R. H. (**1959**). "Masking patterns of tones," J. Acoust. Soc. Am. **31**, 1115–1120.

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

Fastl, H. (**1975**). "Loudness and masking patterns of narrow noise bands," Acustica **33**, 266–271.

Fletcher, H. (**1930**). "A space-time pattern theory of hearing," J. Acoust. Soc. Am. **1**, 311–343.

Fletcher, H. (**1940**). "Auditory patterns," Rev. Mod. Phys. **12**, 47–65.

Florentine, M., and Buus, S. (**1981**). "An excitation-pattern model for intensity discrimination," J. Acoust. Soc. Am. **70**, 1646–1654.

Gässler, G. (**1954**). "Über die Hörschwelle für Schallereignisse mit verschieden breitem Frequenzspektrum," Acustica **4**, 408–414.

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. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hearing Research **47**, 103–138.

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. (**1990**). "Modeling rapid waveform compression on the basilar membrane as multiple-bandpass-nonlinearity filtering," Hear. Res. **49**, 39–60.

Green, D. M. (**1965**). "Masking with two tones," J. Acoust. Soc. Am. **37**, 802–813.

Green, D. M. (**1988**). "Profile Analysis," Oxford University Press, Oxford.

Green, D. M., and Swets, J. A. (**1966**). "Signal Detection Theory and Psychophysics," Wiley, New York.

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

Greenwood, D. D. (**1961b**). "Critical bandwidth and the frequency coordinates of the basilar membrane," J. Acoust. Soc. Am. **33**, 1344–1356.

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.

Hall, J. W., Haggard, M. P., and Fernandes, M. A. (**1984**). "Detection in noise by spectro-temporal pattern analysis," J. Acoust. Soc. Am. **76**, 50–56.

Hall, J. W., and Grose, J. H. (**1990**). "Effects of flanking band proximity, number, and modulation pattern on comodulation masking release," J. Acoust. Soc. Am. **87**, 269–283.

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.

Hartmann, W. M., and Pumplin, J. (**1988**). "Noise power fluctuations and the masking of sine signals," J. Acoust. Soc. Am. **83**, 2277–2289.

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. L. van der, and Kohlrausch, A. (**1994**). "Using an excitation-pattern model to predict auditory masking," Hear. Res. **80**, 38–52.

Heijden, M. L. van der, and Kohlrausch, A. (**1995a**). "The role of envelope fluctuations in spectral masking," J. Acoust. Soc. Am. [accepted for publication]

Heijden, M. L. van der, and Kohlrausch, A. (**1995a**). "The role of distortion products in masking by single bands of noise," J. Acoust. Soc. Am. [submitted for publication]

Helmholtz, H. (**1885**). "On the sensations of Tones," Longmans & Co., New York.[republished in 1954 by Dover Publications, New York]

Humes, L. E., and Jesteadt, W. (**1989**). "Models of the additivity of masking," J. Acoust. Soc. Am. **85**, 1285–1294.

Humes, L. E., and Lee, L. W. (**1992**). "Two experiments on the spectral boundary conditions for nonlinear addition of masking," J. Acoust. Soc. Am. **92**, 2598–2606.

Johnson-Davies, D., and Patterson, R. (**1979**). "Psychophysical tuning curves: Restricting the listening band to the signal region," J. Acoust. Soc. Am. **65**, 765–770.

Johnson-Davies, D. B. (**1981**). "The analogy between psychophysical and physiological tuning curves," Unpublished Ph. D. thesis, University of Cambridge, England.

Kidd, G., Mason, C. R., Brantley, M. A., and Owen, G. A. (**1989**). "Roving-level tone-in-noise detection," J. Acoust. Soc. Am. **86**, 1310–1317.

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.

Kohlrausch, A., and Langhans, A. (**1992**). "Differences in masked thresholds between monaural and diotic conditions: influence of contralateral efferent stimulation?," in *Auditory physiology and perception*, edited by Y. Cazals, L. Demany, and K. Horner (Pergamon Press, Oxford), pp. 571–577.

Langhans, A., and Kohlrausch, A. (**1992a**). "Spectral integration of broadband signals in diotic and dichotic masking experiments," J. Acoust. Soc. Am. **91**, 317–326.

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

Lawson, J. L., and Uhlenbeck, G. E. (**1950**). "Threshold signals," McGraw-Hill, New York.

Leshowitz, B., and Wightman, F. L. (**1971**). "On-frequency masking with continuous sinusoids," J. Acoust. Soc. Am. **49**, 1180–1190.

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

Lutfi, R. A. (**1983a**). "Additivity of simultaneous masking," J. Acoust. Soc. Am. **73**, 262–267.

Lutfi, R. A. (**1983b**). "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 $2f1 - f2$ cubic difference tone with the binaural masking-level difference," J. Acoust. Soc. Am. **69**, 216–222.

Maiwald, D. (**1967**). "Berechnung von Modulationsschwellen mit Hilfe eines Funktionsschemas," Acustica **18**, 193–207.

Moore, B. C. J. (**1985**). "Additivity of simultaneous masking revisited," J. Acoust. Soc. Am. **78**, 488-494.

Moore, B. C. J., and Glasberg, B.R. (**1987a**). "Factors affecting thresholds for sinusoidal signals in narrow-band noise maskers with fluctuating envelopes," J. Acoust. Soc. Am. **71**, 69–79.

Moore, B. C. J., and Glasberg, B.R. (**1987b**). "Formulae describing frequency selectivity as a function of frequency and level, and their use in calculating excitation patterns," Hearing Research **28**, 209–225.

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.

Nelson, D. A., and Fortune, T. W. (**1991**). "High-level psychophysical tuning curves: simultaneous masking by pure tones and 100-Hz-wide noise bands," J. Speech and Hearing Research **34**, 360–373.

Patterson, R. D., and Nimmo-Smith, I. (**1980**). "Off-frequency listening and auditory filter asymmetry," J. Acoust. Soc. Am. **67**, 229–245.

Patterson, R. D. (**1974**). "Auditory filter shape," J. Acoust. Soc. Am. **55**, 802–809.

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

Patterson, R. D., and Moore, B. C. J. (**1986**). "Auditory filters and excitation patterns as representations of frequency resolution," in *Frequency selectivity in hearing*, edited by B. C. J. Moore (Academic Press, Oxford), pp. 123–177.

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

Pickles, J. O. (**1988**). "An introduction to the physiology of hearing," Academic Press, London.

Plomp, R., and Bouman, M. A. (**1959**). "Relation between hearing threshold and duration for tone pulses," J. Acoust. Soc. Am. **31**, 749–758.

Richards, V. M. (**1992**). "The effects of level uncertainty on the detection of a tone added to narrow bands of noise," in *Auditory physiology and perception*, edited by Y. Cazals, L. Demany, and K. Horner (Pergamon Press, Oxford), pp. 337–343.

Rosen, S., and Baker, R. J. (**1994**). "Characteristing auditory filter nonlinearity," Hear. Res. **73**, 231–243.

Schöne, P. (**1977**). "Nichtlinearitäten im Mithörschwellen Tonheitsmuster von Sinustönen," Acustica **37**, 37–44.

Schöne, P. (**1979**). "Mithörschwellen-Tonheitsmuster maskierender Sinustöne," Acustica **43**, 197–204.

Schafer, T. H., Gales, R. S., Shewmaker, C. A. and Thompson, P. O. (**1950**). "The frequency selectivity of the ear as determined by masking experiments," J. Acoust. Soc. Am. **22**, 490–496.

Scharf, B. (**1970**). "Critical bands," in *Foundations of modern auditory theory, Vol. I*, edited by J. V. Tobias (Academic Press, New York), pp. 159–202.

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

Stelmachowicz, P. G., and Jesteadt, W. (**1984**). "Psychophysical tuning curves in normal-hearing listeners: Test reliability and probe level effects," J. Speech and Hear. Res. **27**, 396–402.

Veldhuis, R. N. J., Breeuwer, M., and van der Waal, R. G. (**1989**). "Subbandcoding of digital audio signals," Philips J. Res. **44**, 329–343.

Vogten, L. L. M. (**1978**). "Simultaneous pure-tone masking: The dependence of masking asymmetries on intensity," J. Acoust. Soc. Am. **63**, 1509–1519.

Wegel, R. L., and Lane, C. E. (**1924**). "The auditory masking of one pure tone by another and its probable relation to the dynamics of the inner ear," Phys. Rev. **23**, 266–285.

Wier, C. C., Schlauch, R. S., and Norton, S. J. (**1984**). "The relations among critical ratios, critical bands, and intensity difference limens in man," J. Acoust. Soc. Am. **76**, 1051–1056.

Wright, B. A. (**1992**). "The role of envelope fluctuations in an apparent demonstration of suppression in simultaneous masking," J. Acoust. Soc. Am. **91**, 3436–3442.

Zwicker, E. (**1954**). "Die Verdeckung von Schmalbandgeräuschen durch Sinustöne," Acustica **4**, 415–420.

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

Zwicker, E. (**1956**). "Die elementaren Grundlagen zur Bestimmung der Informationskapazität des Gehörs," Acustica **6**, 365–381.

Zwicker, E. (**1965**). "Temporal effects in simultaneous masking by white-noise bursts," J. Acoust. Soc. Am. **37**, 653–663.

Zwicker, E. (**1970**). "Masking and psychological excitation as consequences of the ear's frequency analysis," in *Frequency analysis and periodicity detection in hearing*, edited by R. Plomp and G. F. Smoorenburg (Springer, Berlin), pp. 376–396.

Zwicker, E. (**1974**). "On a psychoacoustical equivalent of tuning curves," in *Facts and models in hearing*, edited by E. Zwicker and E. Terhardt (Springer, Berlin), pp. 132–141.

Zwicker, E. (**1976**). "Mithörschwellen-Periodenmuster amplitudenmodulierter Töne," Acustica **36**, 113–120.

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

Zwicker, E. (**1980**). "Reversed behaviour of masking at low levels." Audiology **19**, 330–334.

Zwicker, E., and Bubel, C. (**1977**). "Einfluß nichtlinearer Effekte auf die Frequenzselektivität des Gehörs," Acustica **38**, 67–71.

Zwicker, E., and Jaroszewski, A. (**1982**). "Inverse frequency dependence of simultaneous tone-on-tone masking patterns at low levels," J. Acoust. Soc. Am. **71**, 1508–1512.

Summary

This thesis focuses on those instances of masking which can be described as spectral or stationary masking and for which, at a first glance, a straightforward explanation in terms of auditory frequency selectivity is possible. Stated differently, only those situations are considered for which the power spectrum model should do a good job. The power-spectrum model of masking is the quantitative elaboration of a view in which masking is reduced to a combination of frequency selectivity and power detection. The term 'power-spectrum model of masking' as it is used in this text does not refer to one particular model but to the whole of underlying assumptions and views. The first question studied in this thesis is: does such a model really reproduce experimental data adequately and if not, what is missing or wrong?

Chapter 2 evaluates the extent to which auditory masking can be predicted from excitation patterns. Excitation patterns are the patterns of stimulation that acoustic stimuli are supposed to evoke in the inner ear. They play an important role in the formulation of the power-spectrum model of masking. A quantitative model proposed by Glasberg and Moore [Hear. Res. 47, 103–138 (1990)] was used to calculate excitation patterns evoked by stationary sounds. Glasberg and Moore's model gives a detailed quantitative description of the auditory frequency selectivity in terms of auditory filters, the parameters of which were derived from data obtained from so-called notched-noise masking experiments. The whole procedure of deriving auditory filter shapes from masking data is exemplary for the power-spectrum view on masking.

In order to evaluate the model, we tested how well it was able to account for masking data reported in the literature. Masking experiments were simulated by calculating the excitation patterns evoked by the same stimuli that are used the experiments: stimuli containing the masking sound (masker) alone and stimuli containing the masked sound (target) as well as the masker. As a threshold criterion for detectability of the target a difference between the two patterns of 1 dB at any frequency was imposed. This criterion is based on the assumption of power detection mentioned earlier.

For narrow-band-noise masking patterns, the method yields a fairly precise prediction of experimental data. For other conditions, however, systematic deviations between model predictions and data are observed. For instance, the model does not reproduce the typical tip-tail shape of psychoacoustical tuning curves. Furthermore,

the nonlinearities in level dependence are not correctly described, and the model fails to reproduce a realistic two-tone masking curve.

The answer to the first part of the question ('does the model reproduce experimental data adequately') is therefore "no". More precisely, the assumptions of the power-spectrum model of masking are not consistent with all psychoacoustic data considered in this Chapter. In particular, the estimates of level-dependent auditory frequency selectivity derived from notched-noise experiments (Glasberg and Moore, 1990) are in conflict with the following characteristics of masking:

- the large differences in masking effectivity between tones and noise bands
- the amount of upward spread of masking for tonal maskers
- the very shallow growth of masking produced by tonal maskers towards lower frequencies

Furthermore, an analysis of the auditory filters of the model revealed that, in the case of a single narrow-band masker,

- the predicted effect of off-frequency listening on the amount of masking is marginal
- a multi-band detection criterion does not lead to improved prediction of masking.

In the light of these conclusions, the answer to the second part of the question ('what's wrong with the model?') can be transformed to: 'what makes tones and noise bands different maskers?' Once we know this, we will be able to tell why the model fails. Chapters 3 and 4 report experimental work conducted to gain insight into the differences between tonal and noise maskers. These two Chapters deal with two opposite spectral conditions: masking towards frequencies above and below the masker frequency, respectively.

Two experiments are reported in Chapter 3. In the first experiment the masking effect of 5 different types of narrow-band maskers was compared. The masker was either a tone, a narrow-band Gaussian noise, or a multiplied noise obtained by multiplying a sinusoid with a low-pass Gaussian noise. The noise maskers had a bandwidth of either 20 or 100 Hz. In all cases the masker had a center frequency of 1.3 kHz and a duration of 500 ms. Five-point growth-of-masking functions were measured using a 2-kHz tonal target with a duration of 400 ms, temporally centered in the masker.

The data showed several common trends. First, the tonal maskers produced more masking than the noise maskers. Second, Gaussian noise maskers produced more masking than multiplied noise maskers of the same bandwidth. Finally, 100-Hz wide noise maskers produced more masking than 20-Hz wide maskers of the same type. Differences in masked thresholds between the various masker types generally increased with masker level, and exceeded 25 dB in some conditions.

The results indicate that the steep growth of masking which can observed if the masker is a pure tone (the so-called upward spread of masking) is not present when the masker is narrow-band of noise. In that case, a more linear masking growth of masking is observed. Furthermore, it is shown that there are two aspects of noise maskers that determine the seize of this masking release: the statistics of the masker fluctuations and the rate of these fluctuations. The slower the rate of fluctuations (the rate is proportional to the bandwidth), the larger the masking release. This agrees with results obtained with other types of maskers such as two-tone complexes. The role of the masker statistics is more tricky, since the statistics of two types of noise used in the first experiment differ in two aspects, *viz.* fine structure (waveform details) and envelope. A second experiment was performed to find out which of the two aspects is predominant.

In the second experiment the masking effect was investigated for a band-pass noise at 1.3 kHz, with regular zero crossings, but with the envelope characteristics of a 100-Hz wide Gaussian noise. Five-point growth-of-masking functions were measured using a tonal target of 2 kHz. Masked thresholds produced by this hybrid masker were not significantly different from those produced by a 100-Hz Gaussian masker, but differed significantly from those produced by 100-Hz wide multiplied noise. This result indicates that differences in masking between Gaussian and multiplied noise are not due to their different fine structure, but to their different envelope statistics. The largest release from masking occurs when the envelope has marked minima, as is the case for multiplied noise.

The final part of Chapter 3 presents a simple transform-and-average scheme which is shown to account for the observed effect of masker envelope statistics. This scheme regards a fluctuating masker as a concatenation of tonal maskers. A compressive transformation followed by an averaging procedure then reduces the steep growth of masking for tone maskers to the more linear growth of masking for noise bands. It is thus shown that, in order to understand such effects, there is no need for complicated multi-channel models.

Chapter 4 deals with masking towards frequencies below that of the masker. Tonal maskers are known to produce only a very shallow growth of masking under this condition whereas noise bands, again, behave more linearly. In order to explain this more linear behavior of noise, it has been put forward that nonlinear distortions of the ear play a role. In short, it is assumed that the nonlinear interaction of different spectral masker components produces a 'secondary noise.' Particularly below the lower spectral edge of the masker, these distortion products – if they exist – would form a much more effective masker than the original band of noise. If this explanation is correct the growth of masking merely depends on the growth of the intensity of the distortion products with masker level. From what is known about the ear's nonlinear distortions, it is plausible to expect a less shallow, thus more linear growth of masking than for a tonal masker.

Two experiments are reported in Chapter 4. They explore a method for evidencing the role of distortion products as described above. The idea behind the method is the combination of two facts:

- when the same ('frozen') noise sample is used as a masker throughout a masking experiment, the audibility of a tone placed *within* the spectrum of the masker depends on its starting phase.

- when the tone is *outside* the spectrum of noise, masker and tone have no well-defined relative phase.

The first fact is the consequence of the interaction of the tone and nearby noise components; their relative phase determines the increment in energy caused by adding the tone. This leads to an easily measurable (10 dB) effect of signal phase on the masked threshold, as has been known for a long time. The second fact implies that no effect is to be expected for spectrally non-overlapping masker-target conditions. Hence the observation of a phase effect indicates that the masker and the target spectra overlap. This is usually under control of the experimenter, except when the 'effective' masker spectrum is generated by the ear itself, as is supposed to be the case in 'downward masking' by a noise band.

Two masking experiments with frozen-noise maskers were conducted in order 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 a bandpass noise masker ranging from 1 to 2 kHz. In all the 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.

Threshold changes 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 degrees. On increasing the masker level, these sinsuoid-like patterns shifted corresponding to a decreasing phase. The shift for the masker range of 45 dB was between 100 and 200 degrees for the four subjects. The direction of this shift agreed with data on the phase of cubic difference tones as a function of the level of sinusoidal primaries.

The results of this experiment present evidence for the existence and masking potency of distortion products generated by the mutual interaction of noise components. Auditory nonlinearities generated noise components below the spectrum of the original noise. This secondary noise, or 'combinational aggregate' was responsible for the downward masking over a large dynamic range of the masker. The level-induced phase effects in frozen-noise masking showed that the combinational aggregate had characteristics comparable to those of odd-order combination tones such as the cubic difference tone.

All the data of this experiment 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. In a second experiment a large number of different frozen-noise samples were used as maskers in order to evaluate the generalizablity 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 degrees.

An analysis of variance showed that signal phase played a significant role for the high-frequency, but not for the low-frequency masker. Hence the phase effect is not caused by some particularity of the single noise sample used as a masker.

The experimental work presented in Chapters 3 and 4 led to several conclusions with regard to the different masking behavior of noise and tone maskers. The relatively linear masking behavior of noise is a deviation from the 'nonlinear standard' set by tonal maskers, in the sense that it is caused by specific properties of the noise. These specific determinants of noise masking are entirely different for 'upward' and 'downward' masking, *i.e.* masking towards frequencies above and below the masker spectrum, respectively.

Having analyzed differences between tone and noise maskers, we can finally return to the original question: 'what's missing in the power-spectrum model of masking?' The answer is simply that the model incorporates neither of the two aspects of noise maskers that have been found to affect masking profoundly: envelope fluctuations and distortion products. Paradoxically, the model generally describes the noise-masker data better than the tone-masker data. This observation suggests that at least one of the two 'special effects' of noise maskers played a role in the notched-noise masking experiments from which the auditory filter shapes of the model were derived. Indirectly, these observations narrow down the range of applicability of the power-spectrum model of masking. More generally, they undermine the degree of reality that can be attached to auditory filters, for if they cannot be reliably measured then they do not exist.

Samenvatting

Dit proefschrift richt zich op die gevallen van maskering die als spectrale of stationaire maskering kunnen worden omschreven, en die op het eerste gezicht een eenvoudige verklaring in termen van auditieve frekwentieselectiviteit toelaten. Anders gezegd, slechts die gevallen worden beschouwd waarin het vermogensspectrummodel van maskering goed zou moeten werken. Dit model is de kwantitatieve uitwerking van een opvatting waarin maskering wordt teruggebracht tot een combinatie van frekwentieselectiviteit en vermogensdetectie. De term 'vermogensspectrummodel van maskering' zoals in de tekst gebezigd, verwijst niet zozeer naar een bepaald model als wel naar het geheel van achterliggende aannamen en opvattingen. De eerste vraagstelling in dit proefschrift is: reproduceert zo'n model werkelijk de experimentele data op een adequate manier en, zo niet, wat ontbreekt er dan of wat is er fout?

Hoofdstuk 2 evalueert hoe goed auditieve maskering uit excitatiepatronen kan worden voorspeld. Excitatiepatronen zijn de patronen van stimulatie in het binnenoor die akoestische stimuli geacht worden teweeg te brengen. Zij spelen een belangrijke rol in het vermogensspectrummodel van maskering. Een kwantitatief model van Glasberg en Moore [Hear. Res. **47**, 103 138 (1990)] werd gebruikt om de door stationaire geluiden opgewekte excitatiepatronen te berekenen. Glasberg en Moore's model geeft een gedetailleerde beschrijving van de auditieve frekwentieselectiviteit in termen van auditieve filters. De parameters van deze filters werden afgeleid uit zgn. *notched noise* (bandstop-ruis) experimenten.

Dit model werd gebruikt om maskeerexperimenten uit de literatuur te simuleren. Maskeerexperimenten werden gesimuleerd door de excitatiepatronen te berekenen van dezelfde stimuli die in de experimenten gebruikt worden: stimuli die alleen de maskeerder bevatten, en stimuli die behalve de maskeerder ook het testsignaal bevatten. Als drempelcriterium werd een verschil tussen de twee patronen van 1 dB op een willekeurige frekwentie opgelegd. Dit criterium is gebaseerd op bovengenoemde aanname van vermogensdetektie.

In het geval van maskeerpatronen voor smalbandige ruis, leidt de methode tot een redelijk precieze voorspelling van de data. Voor andere condities worden echter systematische afwijkingen tussen modelvoorspellingen en data waargenomen. Bijvoorbeeld, het model slaagt er niet in de typische punt/staart vorm van psychoakoestische *tuning curves* te reproduceren. Daarnaast worden de niet-lineariteiten in maskeerniveau-afhankelijkheid niet correct beschreven, en reproduceert het model ook geen realistische twee-toon maskeercurve.

Het antwoord op het eerste gedeelte van de vraagstelling ('reproduceert het model de experimentele data correct?') is dus "nee". Om precies te zijn, de aannamen achter het model zijn niet consistent met alle psychoakoestische data die in dit Hoofdstuk aan de orde komen. In het bijzonder zijn de schattingen van niveau-afhankelijke auditieve frekwentieselectiviteit zoals afgeleid uit *notched-noise* experimenten (Glasberg en Moore, 1990) strijdig met de volgende kenmerken van maskering:

- de grote verschillen tussen tonen en ruis wat betreft hun maskeer-effectiviteit
- de hoeveelheid opwaartse spreiding bij tonale maskering
- de zeer flauwe groei van maskering van lagere frekwenties door tonale maskeerders

Daarnaast bracht een analyse van de filters van het model aan het licht dat, in het geval van een enkele smalbandige maskeerder,

- het voorspelde effect van *off-frequency listening* op de hoeveelheid maskering verwaarloosbaar is.
- een multiband detectiecriterium niet tot een betere voorspelling van maskering leidt.

In het licht van deze conclusies kan het antwoord op het tweede gedeelte van de vraagstelling ('wat is er fout aan het model?') omgevormd worden tot: 'wat maakt tonen en ruisbanden verschillende maskeerders?' Als we dit eenmaal weten, kunnen we begrijpen waarom het model faalt. Het experimenteel werk beschreven in Hoofdstuk 3 en 4 werd gedaan om inzicht te verkrijgen in de verschillen tussen tonale en ruis-maskeerders. Deze twee Hoofdstukken behandelen twee tegengestelde spectrale condities: resp. maskering naar frekwenties boven en beneden de maskeerfrekwentie.

In Hoofdstuk 3 worden twee experimenten gerapporteerd. In het eerste werd de maskeereffecten van 5 verschillende smalbandige maskeerders vergeleken. De maskeerder was ofwel een toon, ofwel een smalle band Gaussische ruis, ofwel multiplicatie-ruis, verkregen door een sinus te vermenigvuldigen met een laag-doorlaat Gaussische ruis. De ruismaskeerders hadden een bandbreedte van 20 of 100 Hz. De centrumfrekwentie van de maskeerder was steeds 1.3 kHz. Vijfpunts maskeerfuncties werden gemeten met behulp van een testsignaal met een frekwentie van 2 kHz en een duur van 400 ms, temporeel gecentreerd in de maskeerder.

De data vertoonden verschillende trends. Ten eerste bleken de tonale maskeerders meer te maskeren dan de ruisbanden. Ten tweede produceerden de Gaussische ruisbanden meer maskering dan de multiplicatieve ruis van dezelfde bandbreedte. Tenslotte waren de 100 Hz ruisbanden betere maskeerders dan de 20 Hz banden van hetzelfde type. De maskeerverschillen namen i.h.a. toe met toenemend maskeerderniveau en overtroffen in enkele gevallen 25 dB.

De resultaten tonen aan dat, indien een tonale maskeerder, die een steile groei van maskering veroorzaakt, vervangen wordt door een ruismaskeerder, de groei minder steil en dus lineairder wordt. Er wordt bovendien aangetoond dat er twee aspecten zijn die deze afname van maskering bepalen: de statistiek van de maskeerder-fluctuaties en de snelheid van deze fluctuaties. Hoe trager de fluctuaties (de snelheid is evenredig met de bandbreedte), hoe groter de afname van maskering. De rol van de maskeerder statistiek is onduidelijker, omdat de statiek van de twee gebruikte ruizen in twee aspecten verschillen, nl. in fijnstruktuur (golfvormdetails) en omhullende. Een tweede experiment werd uitgevoerd om te bepalen welk van deze aspecten bepalend is.

In het tweede experiment werd de maskering onderzocht door aan ruisband rond 1.3 kHz met regelmatige nuldoorgangen maar met de omhullende karateristieken van een 100 Hz brede Gaussische ruis. Vijfpunts maskeerfuncties werden gemeten m.b.v. een testsignaal op 2 kHz. De drempels bij deze hybride maskeerder bleken niet significant te verschillen van die bij een 100-Hz Gaussische ruis, maar wel van die bij een 100-Hz multiplicatieve ruis. Dit resultaat toont aan dat de verschillen in maskering tussen Gaussische ruis en multiplicatieve ruis veroorzaakt worden door verschillen in omhullende, en niet door verschillen in fijnstruktuur. De grootste afname van maskering treedt op als de omhullende van de maskeerder duidelijke minima vertoont, zoals bij de multiplicatieve ruis het geval is.

In het laatste gedeelte van Hoofdstuk 3 wordt een simpel schaal-en-middeling schema gepresenteerd waarvan wordt aangetoond dat het het geobserveerde effect van maskeerderstatistiek juist beschrijft. In dit schema wordt een fluctuerende maskeerder als een aaneenschakeling van tonale maskeerders beschouwd. Een compressieve transformatie gevolgd door het nemen van een tijdsgemiddelde reduceert de steile groei van maskering voor tonale maskeerders tot de meer lineaire groei voor ruisbanden. Op deze wijze wordt aangetoond dat om zulke effecten te begrijpen, geen beroep gedaan hoeft te worden op ingewikkelde *multi-channel* modellen.

Hoofdstuk 4 behandelt maskering van frekwenties beneden de maskeerfrekwentie. Zoals bekend, vertonen tonale maskeerders slechts een zeer flauwe groei van maskering in deze situatie terwijl ruisbanden, wederom, een lineairder maskeereffect teweeg brengen. Bij wijze van verklaring is de mogelijkheid geopperd, dat niet-lineaire vervormingen van het oor hierbij een rol spelen. Het komt erop neer dat de niet-lineaire interacties tussen de verschillende spectrale componenten van de ruisband een 'secondaire ruis' zouden veroorzaken. In het bijzonder beneden de lage afsnijfrekwentie van de ruisband zullen deze vervormingsprodukten – als ze bestaan – een veel effectievere maskeerder vormen dan de oorspronkelijke ruisband. Indien deze uitleg correct is dan hangt de groei van de maskering voornamelijk af van de groei van de intensiteit van de vervormingsproducten met de intensiteit van de primaire ruisband (de eigenlijk maskeerder). Wat bekend is over niet-lineaire vervormingen in het oor, maakt het aannemelijk een minder flauwe, dus lineairdere groei te verwachten dan voor een tonale maskeerder.

Er worden twee experimenten beschreven in Hoofdstuk 4. Beide verkennen een methode om de bovenbeschreven rol van vervormingsprodukten concreet aan te tonen. De idee achter de methode is de combinatie van twee feiten:

- indien hetzelfde ruisfragment (zgn. *frozen noise*) gedurende het hele experiment als maskeerder wordt gebruikt, hangt de hoorbaarheid van een toon *in het spectrum van de maskeerder* van de *fase* af waarmee deze gestart wordt.
- indien de toon zich *buiten* het maskeerderspectrum bevindt, dan hebben de toon en de ruis geen welgedefinieerde relatieve fase.

Het eerste feit is het gevolg van de interactie van de toon en naburige componenten van de ruis; hun relatieve fase bepaalt de toename in energie die het toevoegen van de toon veroorzaakt. Dit leidt tot een gemakkelijk meetbaar (10 dB) effect van signaalfase op de maskeerdrempel, zoals reeds lang bekend is. Het tweede feit heeft tot gevolg dat er geen fase-effect te verwachten valt als maskeerder en signaal niet spectraal overlappen. Dit al of niet overlappen is gewoonlijk onder controle van de experimentator, behalve wanneer het 'effectieve' maskeerderspectrum door het oor zelf wordt voortgebracht, zoals verondersteld wordt in 'benedenwaartse maskering' door een ruisband.

Twee experimenten met *frozen noise* maskeerders werden verricht om de rol van vervormingsprodukten te onderzoeken, die zouden onstaan door de interactie tussen de componenten van een band-pass ruis. In het eerste experiment werden drempels gemeten van een 900-Hz sinus (duur: 50 ms, incl. 10 ms *ramps*) in de aanwezigheid van een bandpass ruis met afsnijfrekwenties van 1 en 2 kHz. In alle metingen van het eerste experiment werd hetzelfde ruisfragment gebruikt, aangeboden op niveaus van 35, 50, 65 en 80 dB SPL. Het testsignaal werd temporeel gecentreerd in de maskeerder.

De drempel van het testsignaal bleek sterk te variëren als functie van de fase van het signaal; drempelveranderingen van meer dan 10 dB werden gevonden. Het patroon van drempel versus testsignaalfase was sinusvormig; fasewaarden bij minimum en maximum drempel verschilden ongeveer 180°. Een verhoging van het maskeerderniveau resulteerde in het verschuiven van de sinusvormige patronen in de richting van een afnemende fase. Voor de vier proefpersonen varieerde de totale faseverschuiving over het 45 dB masecerernieau tussen 100° en 200°. De richting en orde van grootte van de fase verschuiving kwam overeen met data over de faseverschuiving van de kubische verschiltoon als funktie van het niveau van de primaire tonen.

De resultaten van dit experiment dragen sterke aanwijzingen aan voor het bestaan van de 'secondaire ruisband' en de rol van deze ruisband als werkelijke maskeerder van frekwenties beneden de primaire ruisband. De faseverschuivingen die optraden bij het veranderen van maskeerderniveau suggeren een verwantschap met oneven combinatietonen.

Alle data van het eerste experiment zijn verzameld m.b.v. een enkel ruisfragment van 500 ms. Zo'n aanpak laat een nauwkeurige meting en data-analyse toe, maar heeft het nadeel dat de data enigszins 'anecdotisch' zijn: niets garandeert dat het gekozen ruisfragment in een of ander opzicht speciaal is, en niet de maskerende eigenschappen vertoont van het ensemble waar het uit getrokken is. In een tweede experiment werd een grote hoeveelheid verschillende *frozen-noises* als maskeerders gebruikt. Het doel was uit te zoeken of het fase-effect generaliseerbaar is. Twee soorten ruismaskeerders werden gebruikt: een laagfrekwente (0-800 Hz) variant en een hoogfrekwente (1000-2000 Hz) variant. Voor ieder ruisfragment werd de drempel van een 900 Hz testsignaal gemeten met slechts twee waarden van de signaalfase: 0 en 180 graden.

Een variantie-analyse toonde aan dat het effect van testsignaalfase significant was in het geval van de hoogfrekwente, maar niet in het geval van de laagfrekwente maskeerder. Het fase-effect wordt dus niet veroorzaakt door een of andere eigenaardigheid van het unieke ruisfragment dat in het experiment gebruikt werd.

Het experimentele werk beschreven in Hoofdstukken 3 en 4 geeft aanleiding tot een aantal conclusies m.b.t. de verschillende maskeereffecten van tonen en ruis. Het relatief lineaire maskeergedrag van ruis is een afwijking van de 'niet-lineaire standaard' van de tonale maskeerders, in die zin dat het veroorzaakt wordt door specifieke eigenschappen van de ruis. Deze specifieke determinanten van ruismaskering zijn totaal verschillend voor 'opwaartse' en 'neerwaartse' maskering, d.w.z., maskering van frekwenties boven resp. beneden het spectrum van de maskeerder.

Na deze analyse van de verschillen tussen ruis en tonen als maskeerder keren we terug naar de oorspronkelijke vraag: "wat ontbreekt er aan het vermogensspectrummodel van maskering?" Het antwoord is simpelweg dat het model geen van de twee aspecten van ruismaskeerders incorporeert, die zo bepalend zijn gebleken voor de maskering: omhullende-fluctuaties van de maskeerder en vervormingsproducten. Paradoxaal genoeg beschrijft het model de data van ruismaskeerders beter dan die van maskering door tonen. Dit doet vermoeden dat tenminste één van de 'speciale effecten' van ruismaskeerders een rol hebben gespeeld in de *notched-noise* experimenten waaruit de auditieve filters zijn afgeleid. Indirect beperken deze observaties het toepassingsbereik van het vermogensspectrum van maskering. Meer algemeen ondermijnen ze de graad van realiteit die kan worden toegekend aan auditieve filters, want als ze niet betrouwbaar gemeten kunnen worden, dan bestaan ze niet.

Curriculum vitae

- geboren 11 december 1962 in Tilburg
- '75 – '81: VWO dagopleiding aan het Maurick College in Vught
- juni '81: eindexamen gymnasium β
- '81 '88 studie natuurkunde aan de Rijks Universiteit Utrecht
  - jan '85: kandidaats examen
  - augustus '88: doctoraal natuurkunde;
    afgestudeerd op een onderwerp uit de statistische
    mechanica bij de vakgroep theoretische fysica
- juli '89 – feb '91: vervangende dienstplicht op het Instituut
  voor Perceptie Onderzoek (IPO) in Eindhoven
- mei '91 – mei '95: assistent in opleiding op het IPO

Artikel 21

het is niet toegestaan advertenties in het proefschrift op te nemen.
Artikel 20 van het promotiereglement van de TUE.

Gelukkig mag je wel wat op de titelpagina krabbelen als je de boekjes aan het uitdelen bent – daar ga ik tenminste van uit. Ieder die een exemplaar te pakken heeft gekregen beloof ik bij deze zo'n privé dankwoord op het schutblad. Op eigen risico natuurlijk.

De uitzondering die ik maak, maak ik graag. De wat suffe betiteling 'copromotor' doet geen recht aan Armins enorme betrokkenheid gedurende deze vier jaar. Hij heeft me op sleeptouw genomen, me de duimschroeven aangedraaid wanneer nodig en me ook genoeg ruimte gelaten om er plezier in te houden. Hij heeft me kortom het vak geleerd. Mijn riante status van welbegeleide ajo heeft me veel afgunstige reakties opgeleverd. Dat zegt genoeg.[31]

The surprisingly salient, if not always correlated contributions of aural harmonics (AH) and combination tones (CT) have led me to the valuable insight that even serious distortions can be beneficial, provided they are applied to the right material.

[31] **Deze alinea nu afkappen, anders in strijd met Art. 20.**check refs!!

Stellingen
behorende bij het proefschrift
A comparison of masking by tones and noise
van Marcel van der Heijden

1. Alvorens een verschijnsel kwantitatief te modelleren doet men er goed aan de veronderstellingen die aan zo'n model ten grondslag liggen aan een experimentele test te onderwerpen: voorkomen is beter dan genezen.
   Dit proefschrift, hoofdstuk 2

2. De afleiding van auditieve filtervormen uit maskeerdata zoals gegeven door Glasberg en Moore [Hear. Res. **47**, 103–138 (1990)] is gebaseerd op aannamen die op de door hen gebruikte experimentele condities niet van toepassing zijn.
   Dit proefschrift, hoofdstuk 5

3. In veel gevallen waarin naar Heisenbergs onzekerheidsrelatie wordt verwezen, gebeurt dit te onpas – spreiding is niet hetzelfde als onzekerheid.

4. De centrale rol van eigenwaarden in de gebruikelijke statistische formulering van een quantummechanische meting vormt een onnodige beperking. Een eigenwaardeloos alternatief is het volgende. Zij A een hermitese operator (observabele) en $|\varphi>$ een toestandsvector. Onder zeer algemene voorwaarden, en onafhankelijk van het bestaan van normeerbare eigentoestanden van A, wordt de waarschijnlijkheidsdichtheidsfunctie p_A van meetwaarden van A in toestand $|\varphi>$ geheel vastgelegd door van een welgekozen set testfuncties $\{f_k\}$ de verwachtingswaarden te evalueren:

   $$\int da\, p_A(a)\, f_k(a) = <\varphi \mid f_k(A) \mid \varphi>$$

   Zo levert $f_k : x \mapsto x^k$ het $k$de moment van p_A op en $f_k : x \mapsto e^{ikx}$ de karakteristieke functie.

5. Een ochtend vrij nemen is in het algemeen efficiënter dan een weekend overwerken.

6. Omdat de wiskunde het experiment ontbeert en dus in haar bewijsvoering louter op overtuigingskracht en consensus steunt, is zij onder de exacte wetenschappen de minst objectieve en de meest retorische.

7. Veel apparatuur (*c.q.* software) zou aanmerkelijk aan bedieningsgemak winnen als van de ontwerper geeist werd dat hij, alvorens aan het ontwerp zelf te beginnen, een gedetailleerde handleiding schreef.

8. Een *Theory Of Everything* is net zo'n illusie als perfect viool spelen.

9. De recentelijk ontstane gewoonte om in de lijst met ondertekenaars van overlijdensadvertenties ook de namen van reeds gestorven familieleden op te nemen, is van een macabere absurditeit.

10. Van de welgespelde engelse woorden *psychoacoustics* en *shit* wordt alleen de eerste door de standaard-spellingschecker van UNIX als zodanig erkend. Dit doet vermoeden dat dergelijke databases niet op woordfrequentie gebaseerd zijn.

11. Dat voor wetenschappers geen uitzondering geldt op het aloude 'de beste stuurlui staan aan wal', kan behalve uit de gemiddelde academische koffiepauze opgemaakt worden uit verhandelingen van buiten hun vakgebied geraakte auteurs als Penrose en Sheldrake.