Citation for published version (APA):

DOI:
10.1016/0378-5955(94)90007-8

Document status and date:
Published: 01/01/1994

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

Please check the document version of this publication:
• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license above, please follow below link for the End User Agreement:
www.tue.nl/taverne

Take down policy
If you believe that this document breaches copyright please contact us at:
openaccess@tue.nl
providing details and we will investigate your claim.
Using an excitation-pattern model to predict auditory masking

Marcel van der Heijden, Armin Kohlrausch *

Institute for Perception Research (IPO), P.O. Box 513, 5600 MB Eindhoven, the Netherlands

Received 5 November 1993; revision received 19 June 1994; accepted 23 June 1994

Abstract

This paper evaluates the extent to which auditory masking can be reliably 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 psychophysical tuning curves. Furthermore, the nonlinearities in level dependence are not correctly described, and the model fails to reproduce a realistic two-tone masking curve.

Keywords: Excitation pattern; Model; Masking; Auditory filters; RoEx filters

1. Introduction

A fundamental observation in auditory masking is the importance of the spectral content of the signals. Generally, signals that are spectrally alike, tend to mask each other most strongly. For instance, the masking pattern produced by a narrow band of noise shows a clear maximum at the center frequency of the masker (Egan and Hake, 1950). If controls are used to minimize the unwanted influence of temporal factors, such as beats, then spectral aspects of masking can contribute to our understanding of the mechanisms underlying the frequency selectivity of the auditory system (Patterson and Moore, 1986).

Spectral aspects of masking are generally believed to reflect the spread of excitation along the tonotopically organised 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 fibres 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 (Fletcher, 1940; Patterson,
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 signal.

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. 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 (1987) and Glasberg and Moore (1990). With the help of these filter shapes, excitation patterns of arbitrary signals can be calculated. Moore and Glasberg suggest their use in evaluating different aspects of hearing, including masking of arbitrary stationary maskers (Moore and Glasberg, 1987, 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 Steemerdink, 1992; Brandenburg and Sporer, 1992).

2. General method

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

![Schematic diagram of the calculation of excitation patterns.](image-url)

Fig. 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.
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 signal 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).

Fig. 1 shows the building blocks of the model. Excitation patterns are calculated from the amplitudes of the discrete signal 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 signal 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 signal level. With increasing input level the lower slope of the filter becomes shallower. The contribution of each signal 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 signal level within the one-ERB band surrounding the signal component under consideration.

The filter bank splits the signal 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. 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 signal. The spectral content of this noise is chosen so as to reproduce the absolute threshold for tonal signals. 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.

In Fig. 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 signal 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 (Schiöne, 1979; see also Fig. 7 below). Apart from the limitation at lower levels caused by the internal noise, peaks and lower skirts of the patterns grow linearly with level.

2.2. Procedure

Given the procedure to calculate the excitation pattern evoked by a spectrally-represented signal, 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. 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

---

1 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.

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

---

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).
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) \(^4\). A more complete quantitative description of OFL is given in the Appendix.

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 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 Appendix it is 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 (Gässl er, 1954; Langhans and Kohlrausch, 1992). 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, namely 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.

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. 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 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.

---

\(^4\) 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).
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.

3. Simulations

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.

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.

In Fig. 4, 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 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. 5 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 masker 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.

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.

---

5 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.
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. 5) also predicts 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.

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. 6, 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. 4 and 5), 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. 6). 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.

### 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. 7 together with Schöne’s data. Data for identical masker and target frequencies (Δz = 0) are included in both panels.

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. 7). 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. 7) 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 Δz, ranging from 0.95 dB/dB at \( Δz = 0 \) Bark to 2.1 dB/dB at \( Δz = 4 \) 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. 7) 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. 7).

The growth-of-masking functions of Bacon and Viemeister (1985), obtained with short (20 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 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. 8 (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.

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 vs. 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 mask-

### Table 1

<table>
<thead>
<tr>
<th>Masker frequency (Hz)</th>
<th>Measured slope of growth-of-masking function</th>
<th>Standard deviation</th>
<th>Calculated slope of growth-of-masking function</th>
</tr>
</thead>
<tbody>
<tr>
<td>690</td>
<td>1.55</td>
<td>0.26</td>
<td>1.43</td>
</tr>
<tr>
<td>800</td>
<td>1.28</td>
<td>0.04</td>
<td>1.27</td>
</tr>
<tr>
<td>909</td>
<td>0.93</td>
<td>0.05</td>
<td>1.07</td>
</tr>
<tr>
<td>1000</td>
<td>0.89</td>
<td>0.05</td>
<td>0.93</td>
</tr>
<tr>
<td>1052</td>
<td>0.67</td>
<td>0.08</td>
<td>0.95</td>
</tr>
<tr>
<td>1111</td>
<td>0.59</td>
<td>0.07</td>
<td>0.86</td>
</tr>
<tr>
<td>1250</td>
<td>0.26</td>
<td>0.07</td>
<td>0.75</td>
</tr>
</tbody>
</table>

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.
ing in terms of spectral resolution alone is an oversim-
plification. Temporal effects that evolve from masker
fluctuations could play a role in the observed differ-
eence. 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.

3.3. Psychophysical tuning curves

Unrestricted tuning curves

A psychophysical 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 fre-
quency, 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 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 physi-
ological 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 0.4 \( f_T \) to 1.2 \( f_T \), with a
duration of 500 ms. The targets had a duration of 250
ms and were temporally centered in the maskers. The

![Fig. 8. NBN-on-tone growth-of-masking function for a target-masker
distance of 2.6 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.](image)

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 For-
tune (1991) are reproduced in Fig. 9 (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 fre-
quency 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 to-
wards lower frequencies is broadened; this is an indica-
tion 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 condi-
tions include usage of short (50 ms) targets (Vogten,
The solid lines in Fig. 9 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.

Because of the great 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 method section). 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 'multiband 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 calculated. 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 non-integrative 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. 9: 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. 9, respectively), correspond to intensity JNDS. For the low-level PTCs the multiband model predicts a lower masker level at threshold than 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 show less masking at higher levels. 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 Appendix). 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).

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. 10 together with the data.
of Johnson-Davies and Patterson. Open triangles in all panels in Fig. 10 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. 9; 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 of Fig. 10 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 Appendix). This trend is also seen in the data; adding the fixed masker decreases the upper slope of the PTC. In the 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. 10 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. 10) 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. 10. 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. 10). 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 the Appendix, 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. 10, 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. 10, 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 the Appendix, 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. 10 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).

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 $\Delta f$ with the critical bandwidth at the target frequency. A newer set of data is presented in Zwicker (1980), from which the data in Fig. 11 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 $\Delta f$ 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 large $\Delta f$, 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. 12 shows the $\Delta f_{2,t}$ 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 $\Delta f_{-3dB}$ of the same RoEx filter (open triangles). A comparison between the curves shows that, when applied to the 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 $\Delta f_{2,t}$ 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, 1965, Florentine and Buus, 1981). In any case, the results shown in Fig. 12 warn against too simple an interpretation of experimental data in terms of theoretical concepts.

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 underlie 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, 1984) 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 Appendix).

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 bandfilter approach, which lacks suppression and other nonlinear behavior. These deficiencies 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 energy 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 energy detection fails to explain the observed phenomena (Kidd, 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 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 energy 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. the section on growth-of-masking functions). 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.

Acknowledgement

We would like to thank D.J. Hermes, A.J.M. Houtsma and B.C.J. Moore for many useful comments on an earlier version of this paper.

Appendix: Off-frequency listening and masking

In a simulation of a simple tone-on-tone masking condition (cf. Fig. 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 Appendix 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 signal (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. A1 shows an example of the excitation patterns under consideration. It shows patterns 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 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 signal 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. A2. 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 general method 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. A2 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. A3 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

---

6 Since masking is determined by the simultaneous excitation of T and M, it is necessary to adjust the filter shapes according to the overall 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.
The lower panel shows the effect of combining the two maskers (solid lines). Detection is limited to the valley between the maskers, the frequencies of optimal detection are indicated by the arrows, yielding a target threshold of 35 dB (dotted line). The dotted line is the excitation pattern of the target at threshold alone. The dashed lines are the patterns of two sinusoidal maskers at 700 and 1300 Hz. The excitation patterns of two sinusoidal maskers at 700 and 1300 Hz. In the upper panel, the solid lines indicate detectability in the single-masker case is spoiled by 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. A3), 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 signal frequency (arrow in the lower panel of Fig. A3). 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.

Fig. A3. 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).

patterns for low masker + target and high masker + target are given by the dashed lines in Fig. A3. 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. A3), 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 signal frequency (arrow in the lower panel of Fig. A3). 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.

References


Fig. A3. 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).


