Acta Acustica united with Acustica
Vol. 88 (2002) 348 – 358

Linear and Nonlinear Processes in Temporal Masking

Christopher J. Plack
Department of Psychology, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ, UK

Andrew J. Oxenham
Research Laboratory of Electronics, Bldg. 36-763, Massachusetts Institute of Technology, Cambridge, MA 02139-4307, USA

Vit Drga
Department of Psychology, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ, UK

Summary
A number of masking phenomena can be modeled in terms of a linear auditory filter bank followed by a temporal integrator and a simple decision device based on the signal-to-masker ratio. Other aspects require the inclusion of a nonlinearity following linear filtering. The present article concentrates on aspects of non-simultaneous, or "temporal", masking that cannot be explained by either modeling approach. A more realistic nonlinear model of cochlear processing is explored, and it is shown that most aspects of temporal masking can be accounted for with this model, in conjunction with a linear temporal integrator (temporal window) and the simple decision criterion. An earlier version of the model with a static nonlinearity following linear filtering can account for "on-frequency" nonlinearities in masking such as the growth of forward masking, the effects of forward masker duration, and the nonlinear additivity of forward and backward masking. The new model can account, in addition, for frequency-dependent nonlinearities in masking such as the upward spread of forward masking and certain aspects of two-tone suppression.

PACS no. 43.64.Kc, 43.66.Ba, 43.66.Dc

1. Introduction

It used to be common, as a matter of convenience, for models of masking to assume that the filtering processes in the cochlea are linear. Specifically, the cochlea has often been modeled as a bank of linear band-pass filters [1, 2, 3]. Furthermore, it is often assumed in these models that the ratio of signal intensity to masker intensity at threshold is invariant with level. Although it has been known for some time that the frequency response of the auditory system is nonlinear [4], these simplifying assumptions seem to work well in certain types of masking situations. For example, the signal-to-masker ratio for detecting a tone in a broadband noise masker is roughly invariant with level [5], as is the Weber fraction for the intensity discrimination of broadband noise [6]. Although the Weber fraction for pure tones decreases with level [7], this effect can be explained if it is assumed that the "spread of excitation" along the basilar membrane (BM) as the tone level is raised provides additional information that aids discrimination ([8, 9, 10, 11] but see [12] for an alternative view).

Received 30 August 2001, revised 18 February 2002, accepted 4 March 2002.

That masking in the healthy ear is an inherently *nonlinear* process is particularly apparent when the signal and the masker are presented non-simultaneously ("temporal" masking). The present article begins by reviewing the empirical evidence for the hypothesis that nonlinearities encountered in temporal masking are primarily of cochlear origin, and that subsequent temporal processing can be treated as linear in the intensity domain (i.e., linear with respect to the square of BM vibration). New modeling results are then presented that provide quantitative support for this view. Models incorporating a nonlinear stage have been applied with success to many aspects of auditory perception, such as frequency selectivity [13], modulation detection [3], and loudness [14]. In the present article, a nonlinear approach is used to account for aspects of temporal masking.

1.1. The basilar membrane response and temporal masking

Direct physiological measurements of BM vibration [16, 17, 18, 19, 15] have revealed a response that is highly nonlinear. As an illustration, Figure 1 shows data replotted from Ruggero *et al.* [15] recorded from a single place (with a single characteristic frequency, or CF, of 10 kHz) on the BM of the chinchilla. (Following the usual usage

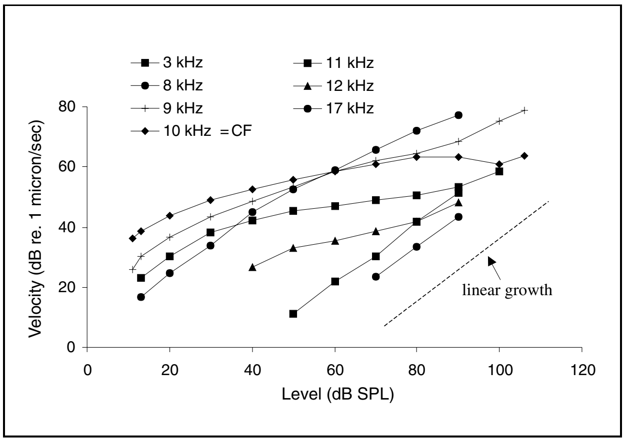

Figure 1. Basilar menbrane velocity-intensity functions for a chinchilla, showing BM velocity (in dB) as a function of input level (in dB SPL) for tones of different frequencies. Velocities were measured at a single place on the BM with a CF of 10 kHz (after Ruggero *et al.* [15], Figure 6).

in the physiological literature, "CF" will be used in this article to refer to the frequency that produces the largest response for a low-level signal.) The curves show BM velocity (in dB) against input level (in dB SPL). Any straight line with a slope of one on these co-ordinates would indicate a linear response. From the point of view of the present article, three main features of the BM response are apparent:

1. The response to a tone at CF is almost linear at low levels.
2. The response to a tone at CF is highly compressive at medium to high levels.
3. The response to a tone well below CF is linear at all levels.

It will be argued in this section that the characteristic features of the BM response can account for all the major nonlinear properties of temporal masking. All that needs to be assumed is that, for a given set of temporal and spectral properties, masked threshold corresponds to a constant ratio of signal excitation to masker excitation following nonlinear cochlear processing. Although it is probable that cochlear processing is much more complex than the account provided here, these main features of the BM response seem to be all that is required to explain a wide variety of psychophysical results.

In forward masking (in which the masker precedes the signal), a given increase in masker level usually causes a much smaller increase in the signal level at threshold, so long as the signal level is less than around 40 dB SPL [20, 21, 22]. The shallow growth for low signal levels can be explained on the basis first, that the response of the BM to a tone at CF is linear at low levels and compressive at higher levels; and second, that the masker level is generally much higher than the signal level at forward-masked threshold. In other words, the masker and signal do not experience the same amount of compression. If the signal is in the low-level, linear region, then a given increase in signal level (in dB) will produce the same increase in BM vibration (in dB). If the masker is at a higher level, and is therefore in the compressive region of the BM function, then a given increase in masker level will produce a smaller increase in BM vibration. It follows that if the signal-to-masker ratio in terms of BM vibration is to remain constant, physical masker level has to grow more rapidly than physical signal level, leading to a shallow slope.

The effect of masker duration on forward masking also varies with level [23, 24, 25], implying a nonlinear process [26]. Specifically, increases in masker duration produce a greater increase in signal threshold when the *signal* is at a high level than when it is at a low level. Related to this is the finding that equally-effective forward and backward maskers do not combine in a linear way for moderate signal levels [27, 28, 29]. The combined masking is usually much greater than the 3-dB increase over the single-masker threshold expected on the basis of linear intensity integration (a 3-dB increase is a doubling in intensity). Both these effects can be explained in terms of the BM response to a tone at CF. Assume that a given change in masker duration produces the same internal effect on masking (i.e., that the integration process *per se* is linear). Now if the signal is in the low-level, linear region of the BM response, then the *physical* signal level at threshold will change by a certain amount to produce the necessary change in BM vibration. However, if the signal is in the compressive region of the BM response, the physical signal level will have to change by a greater amount to produce the same change in BM vibration. This explains the level-dependence of the duration effect. Similarly, if the signal is compressed, then the 3-dB increase in effective masking produced by combining two equally-effective maskers will require a larger than 3-dB increase in physical signal level [30].

Upward spread of masking occurs when the masker is at a lower frequency than the signal. A given increase in masker level is found to produce a greater increase in the signal level at threshold [4, 31, 32, 33]. Upward spread of masking is observed in simultaneous masking, but the largest effect (in terms of slope) is seen in forward masking [34]. It can be seen in Figure 1 that the BM response varies with frequency at a given place in the cochlea. The differential effect of frequency can explain the upward spread of forward masking [34, 35]. If the masker is much lower in frequency than the signal, then the growth in BM vibration with masker level *at the place tuned to the signal* will be roughly linear. However, for moderate levels the response to the signal will be highly compressive. This means that a given increase in masker level will require a greater increase in signal level to maintain a constant signal-to-masker ratio in the BM response. For low signal levels the BM is more linear, and consequently the upward spread of masking is not observed in forward masking [34, 35]. The upward spread of masking is observed at low signal levels in simultaneous masking, however, and it is probable that this is a consequence of suppression [31, 35, 36].

Suppression is one of the most obvious forms of nonlinearity observed in conjunction with temporal masking [37, 38, 39, 40, 41]. Suppression refers to the reduction in the effective level of one sound component in the presence of another component. Houtgast [39] used both forward masking and “pulsation threshold” to demonstrate suppression. In these experiments, suppression is revealed as a reduction in masking when a suppressor is gated with the masker.

Psychophysical tuning curves (PTCs) are plots of the level of a masker required to mask a fixed low-level signal as a function of masker frequency. PTCs tend to be sharper (reflecting greater frequency selectivity) when measured using forward masking than when measured using simultaneous masking [39, 42, 43, 44]. This difference may be explained in terms of suppression if it is assumed that the process that produces the threshold elevation in simultaneous masking is partly excitatory and partly suppressive [35, 36, 45]. Because the suppressive region extends beyond the tuning curve either side of the test frequency [39, 41], the extra effect of suppression in reducing the signal level may reduce the level of the masker needed in these regions, thereby broadening the tuning curve.

Suppression is hard to explain as a simple consequence of the response of the BM to pure tones of different frequencies. Suppression will be explored in more detail in Section 2.3. Suffice to say at this stage that the characteristics of suppression measured psychophysically by Houtgast [39] and by Duifhuis [37] are similar to the characteristics of suppression measured in the auditory nerve [46, 47], although saturation in the auditory nerve leads to some discrepancies [37]. If it is assumed that physiological suppression is the result of nonlinear interactions in the cochlea, then it is reasonable to infer that psychophysical suppression is the result of the same underlying process.

The results described so far apply to listeners with normal hearing. For these listeners it appears that the main nonlinear characteristics of temporal masking (effects of masker level, effects of masker duration, nonadditivity, upward spread of masking, and suppression) are a consequence of the nonlinear response of the cochlea. Further weight is added to this claim when the results from listeners with sensorineural hearing loss are considered. Physiological models have indicated that sensorineural hearing loss is associated with a linearization of the BM response [15, 48]. That is, animals with a moderate to severe artificiallyinduced cochlear hearing loss tend to show a linear BM response to a tone at CF. If the hypothesis regarding the origin of the nonlinearities in temporal masking is correct, then listeners with moderate sensorineural hearing loss should display linear temporal masking characteristics. In fact, this is exactly what has been found.

First, the growth of forward masking with masker level is roughly linear in impaired listeners [28, 49]. Second, hearing-impaired listeners show a reduced effect of masker duration [25], and roughly linear additivity of forward and backward masking [28]. Third, hearing-impaired listeners show a linear growth of masking (i.e., no expansive upward spread of masking) when the masker is lower in frequency than the signal [34, 50, 51]. Finally, suppression is reduced or absent in impaired listeners [52, 53].

1.2. Linear intensity integration and the temporal-window model

The findings outlined in Section 1.1 offer compelling evidence that the BM is the main source of nonlinearity in temporal masking. Furthermore, other measures of temporal resolution, such as gap detection and the temporal modulation transfer function, are largely invariant with level [54, 55, 56]. A popular model of temporal resolution includes an auditory filter, a means of extracting the envelope of the sound, and a lowpass filter or leaky integrator [57, 58]. One version of this approach that has been applied to temporal masking in particular is the temporal-window model [58, 59, 60]. An advantage of these models, which are based on temporal integration, is that they can account for temporal masking and for other temporal resolution phenomena such as gap detection.

The findings described in the previous section suggest strongly that if such an integration process exists, then it may be regarded as *linear in the intensity domain* for modeling purposes. In other words, the integration process itself is a simple weighted summation of a quantity that is a linear function of the intensity of BM vibration. It is interesting to note that firing rate in the auditory nerve may be linearly related to the intensity of BM vibration, at least at levels close to the thresholds of the high spontaneous rate fibres [61, 62, 63]. A square law may be “an accurate description of the underlying synaptic drive to all primary auditory nerve fibres” [61]. Furthermore, there is evidence that the relationship of BM velocity to inner hair cell receptor potential is also a square law [64]. It should be emphasized that the integration process does not appear to be linear in terms of BM displacement or velocity. If it were, then two equally effective maskers should combine to produce 6 dB of additional masking in impaired listeners, and at low signal levels in normally-hearing listeners. This is not observed.

The original temporal-window model integrated *stimulus* intensity. This model did not provide a very convincing account of temporal masking because the shape of the window had to change with level to account for the nonlinear growth [58, 60]. A later incarnation used a nonlinear function before the temporal window to simulate the nonlinear response of the BM to a tone at CF [26, 49, 59, 65]. This model can account for “on-frequency” nonlinearities in temporal masking such as the growth of forward masking [65], the effect of masker duration in forward masking [26], and the nonlinear additivity of forward and backward masking [59]. However, a model such as this cannot account for the *frequency-dependent* nonlinearities in temporal masking, such as upward spread of masking and suppression. To account for these phenomena a proper nonlinear filter must be used before the temporal window. In the next Section, a first attempt at incorporating such a filter is presented.

2. The DRNL temporal-window model

The temporal-window model is a model of temporal masking and of temporal resolution [58, 59, 60, 66]. The model assumes that the internal representation of a stimulus is smoothed in time by the action of a sliding intensity integrator, or temporal window. The temporal window behaves as a lowpass filter acting in the intensity domain. An assumption of the model is that temporal masking, both forward and backward, is a result of this integration process: The signal is masked because the temporal window centred on the time of signal presentation integrates masker energy for times before and after. Forward masking is assumed to arise from a persistence of the neural activity produced by the masker, not because of the reduction in sensitivity associated with neural adaptation [67].

2.1. Incorporating a nonlinear auditory filter

In the version of the model introduced here, the function representing the response of the BM to a tone at CF has been replaced by a nonlinear auditory filter, which shows a compressive response to a tone at CF but a *linear* response to a tone lower than CF. The nonlinear filter used here is a version of the dual-resonance nonlinear (DRNL) filter developed by Meddis and colleagues [68, 69]. The DRNL filter was designed to simulate physiological measurements of the BM response. An earlier filter design that works on similar principals is the multiple-bandpass-nonlinearity (MBPNL) model of Goldstein [70, 71]. However, the DRNL filter has been optimized to more recent physiological data. The compression exponent used in the MBPNL model (generally 0.5) is less than the values reported recently of around 0.2 [15, 49, 72], and (unlike the DRNL filter) the MBPNL filter does not produce the observed intensity-dependent change in best frequency. This can have a large effect on the variation of masked threshold with masker or signal frequency when the level of the masker is high and the level of the signal is low [73], although for the data presented in this article it is not necessary to model the intensity-dependent shift. One advantage of the MBPNL model is that it correctly models the different growth of suppression for low-side and high-side suppressors, as discussed in Section 2.4.

Figure 2 shows a schematic of the different components of the DRNL filter. The components of the whole model are illustrated in Figure 3, and the parameters of the model are presented in Table I. A full description of the model will not be attempted here. The reader is referred to the article by Meddis *et al.* [68] for a description of the basic properties of the DRNL stage. It should be noted that the parameters of the DRNL presented in Table I are different from those used by Meddis *et al.* In particular, the compression exponent used here is 0.16, as compared to 0.25 in the Meddis *et al.* simulation. The nonlinearity in the present study was the same as that used by Plack and Oxenham [65] and by Oxenham and Plack [26], and is based on the masking data of Oxenham and Plack [34]. As illustrated in the latter article, the compression exponent is consistent with physiological measurements in guinea pig and chinchilla [16, 15, 72]. The bandwidths of the gammatone filters used in the present simulation were broadly similar to those used by Meddis *et al.* to simulate the guinea pig recordings of Nuttall and Dolan [74]. The articles by Plack and Oxenham [65] and Oxenham and Plack [26] provide a detailed description of the temporal window stage. The temporal window function used here is identical to that used in the two earlier articles.

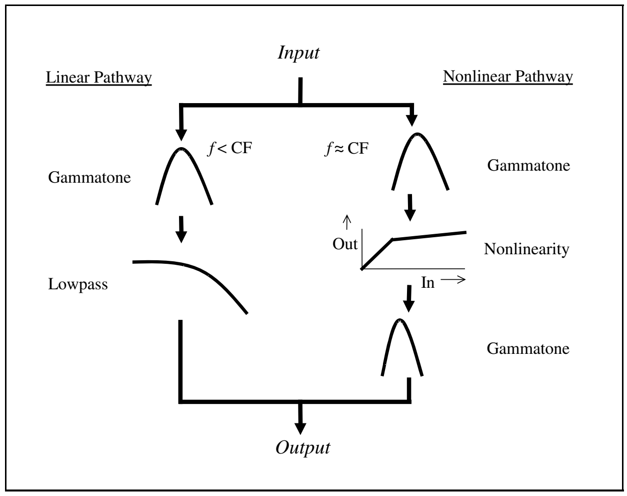

Figure 2. Schematic of the dual-resonance nonlinear (DRNL) filter [68]. The linear pathway consists of a gammatone filter followed by a low-pass filter. The non-linear pathway consists of a gammatone filter, a compressive nonlinearity and a second gammatone filter. Input is processed in parallel through each pathway before being recombined additively.

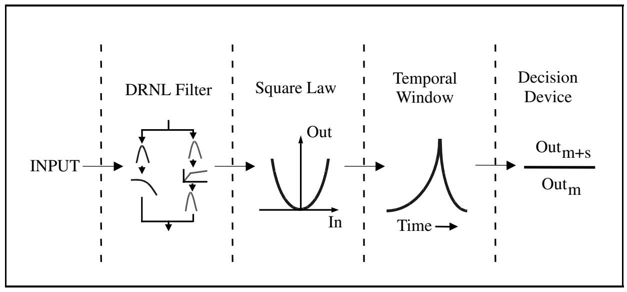

Figure 3. Schematic of the DRNL temporal window model, comprised of the DRNL filter, a square-law rectifier, an exponentially-weighted temporal window / integrator, and a decision device.

In the DRNL stage, the input signal is passed through two parallel pathways. In the nonlinear pathway, the signal is passed through a 2nd-order gammatone filter [75], followed by a nonlinear function similar to that used by Plack and Oxenham [65] in their model of the growth of forward masking. The function is almost linear at low levels and compressive at medium to high levels. The final stages of the nonlinear pathway are another 2nd-order gammatone filter and a linear amplification (or gain). Notice that the nonlinearity is sandwiched between two gammatone filters. As will be seen in Section 2.3, this arrangement is crucial for the simulation of two-tone suppression. Also

Table I. The model parameters used for the simulations. With the exception of the decision variable, k, the same set of parameters were used for all the simulations presented here. For a more detailed description of the meaning of these parameters, see Plack and Oxenham [65] and Meddis *et al.* [68].

Parameter	Value
Lower slope of nonlinearity (dB/dB)	0.78
Upper slope of nonlinearity (dB/dB)	0.16
Transition level between lower and upper (dB SPL)	40
Gain of nonlinear pathway re. linear pathway (dB)	39
1st nonlinear GT filter centre frequency (re. CF)	1.10
1st nonlinear GT filter bandwidth (ERB, re. CF)	0.14
2nd nonlinear GT filter centre frequency (re. CF)	0.94
2nd nonlinear GT filter bandwidth (ERB, re. CF)	0.08
Linear GT filter center frequency (re. CF)	0.93
Linear GT filter bandwidth (ERB, re. CF)	0.12
Linear lowpass filter cutoff frequency (re. CF)	0.91
Linear lowpass filter order	64

notice that the centre frequency of the first gammatone filter is above the CF of the DRNL filter, and that the centre frequency of the second gammatone filter is below the CF. This is another feature of the present model that is not seen in the version presented by Meddis *et al.*, where the two filters in the nonlinear pathway have the same centre frequencies. It will be shown that the higher centre frequency of the first gammatone filter is necessary to reproduce the suppression patterns measured by Houtgast [39].

The linear pathway involves a 2nd-order gammatone filter followed by a low-pass filter. Both the centre frequency of the gammatone and the cut-off frequency of the low-pass filter are lower than CF. The overall effect of the two pathways is that input frequencies close to CF (indicated by $f \approx$CF) pass through the non-linear pathway, and that input frequencies much less than CF (indicated by $f <$CF) pass through the linear pathway. The outputs of the nonlinear pathway and the linear pathway are added. The combined output is squared and then smoothed by the temporal window. When the signal level is low it is sometimes necessary to simulate absolute threshold. For this purpose a constant is added to the output of the temporal window. The constant is determined by the measured absolute threshold for the data set concerned [65].

The output of the temporal window as a function of time is called a "temporal excitation pattern" (TEP) by analogy with the "excitation pattern" in the frequency domain, which represents the output of a bank of auditory filters [1]. The output of the temporal window as a function of CF and time has been called a "spectro-temporal excitation pattern" (STEP [76]) as it reflects the overall spread of excitation across CF and time. The STEP can be regarded as a temporal extension of the standard excitation pattern.

To find predictions for a set of temporal masking data, the masker alone is passed through the model and then the masker plus signal is passed through the model. Detection is assumed to be based on the ratio of the signal-plus-masker at the output of the model to the masker alone at the output of the model, expressed in dB (in other words, a measure of ΔL), at the time that this ratio reaches its maximum value. To reduce the computational load, for forward masking the "centre time" of the temporal window is assumed to be at the start of the offset ramp of the signal. Preliminary work showed that the optimum time for the temporal window in forward masking is close to the beginning of the signal offset ramp, and the predictions obtained by allowing the temporal window to vary in time and choosing the optimum placement do not vary significantly from those obtained by fixing the window placement in this way. In addition, the DRNL filter always had a fixed CF equal to the frequency of the signal. In other words, the parameters in Table I, describing the filter centre frequencies and bandwidths, are always specified relative to the signal frequency in the simulations. Only the output of one auditory filter was considered for each signal frequency.

One of the parameters of the model is k, the value of the decision variable at threshold. In predicting a single data point, the model varies the dependent variable (e.g. signal threshold), repeating the simulation until the value of the decision variable is equal to k. The value of the dependent variable required is taken as the prediction of the model. This process is repeated for the entire data set. For each value of k a constant is added to the output of the temporal window so that, in the absence of the masker, the prediction of the model is equal to the absolute threshold determined experimentally. In the simulations that follow, the only parameter in the model that was varied to fit the data was k. However, for the data presented in each figure, a single value of k was used for all the data points. In other words, for each set of data the model was identical. Allowing k to vary between data sets seems reasonable. Different sets of listeners may be expected to differ in their overall sensitivities, and procedural differences may also result in some variation in performance.

2.2. Simulating upward spread of masking

The model incorporates a nonlinear auditory filter followed by linear intensity integration. It has been shown that the temporal window model with a similar response to a tone at CF as the present model can simulate the "on-frequency" nonlinear characteristics of temporal masking [26, 59, 65]. The new model contributes the ability to simulate the nonlinear characteristics of temporal masking that are related to the frequency-specificity of the nonlinearity.

To illustrate this property of the model, Figure 4 shows the upward spread of masking data of Oxenham and Plack [34]. In this experiment, the 4-ms pure tone signal was presented 2 ms after a 104-ms pure-tone masker. The masker level needed to mask the signal is plotted as a function of signal level. When the masker is at the same frequency as the signal the response grows linearly. Since the masker and the signal are at roughly the same level, either both are passed linearly by the BM at low levels, or both are compressed at moderate levels. Either way, the growth in the BM response with level is the same for signal and masker. When the masker is an octave below the signal frequency, the growth is shallow, because only the signal is being compressed.

The dashed lines in Figure 4 show the predictions of the DRNL temporal-window model. The absolute threshold for the signal, used to calculate the constant added to the output of the temporal window, was 29.6 dB SPL. To find the best fit to the data, the model parameter k was varied adaptively. The best fitting value was 1.41 dB. The model captures the data very well. Because the model includes a linear low-frequency pathway in addition to a nonlinear on-frequency pathway, it is able to account for the differential effects of masker frequency. In the simulations, the CF of the DRNL filter (the best frequency at low levels) was constant. However, because the best frequency of the DRNL filter changes with level (reproducing the so-called “basalward shift” in the traveling wave) the signal was not always at the best frequency of the filter. An inspection of the model confirmed that the shift only becomes significant above about 80 dB SPL, when the linear pathway begins to dominate over the nonlinear pathway. Interestingly, this is also the signal level at which the masking function for the 3-kHz masker becomes more linear in the experimental data. It is possible, therefore, that fixing the filter best frequency rather than CF would have provided an even better fit to the data.

2.3. Simulating two-tone suppression

The classic psychophysical suppression results are those of Houtgast [39]. Although his measurements were made using the pulsation threshold technique, it is possible to simulate the overall form of his data by finding the level and frequency of a suppressor tone required to produce a criterion reduction in the output of the model in response to a masker tone at CF. Houtgast used a 1-kHz masker tone presented at 40 dB SPL. Unfortunately, it is not clear how this translates to dB SPL. However, since the IS0 226 standard absolute threshold (MAF) at 1 kHz is 4.5 dB SPL [77], it seems reasonable to assume that 40 dB SL is not grossly different from 40 dB SPL.

The solid lines in Figure 5 show the boundaries of the suppression regions measured by Houtgast. A suppressor with a level and frequency within the boundary regions produced at least a 3-dB reduction in the level of the test tone needed for pulsation threshold. In other words, the suppressor reduced the effective level of the masker tone by 3 dB. A similar region was determined using just the DRNL stage of the model. A continuous 1-kHz masker tone with a level of 40 dB SPL was passed through the filter with a CF of 1 kHz and the output level in dB was measured. The procedure was then repeated with a suppressor tone of variable frequency and level added to the

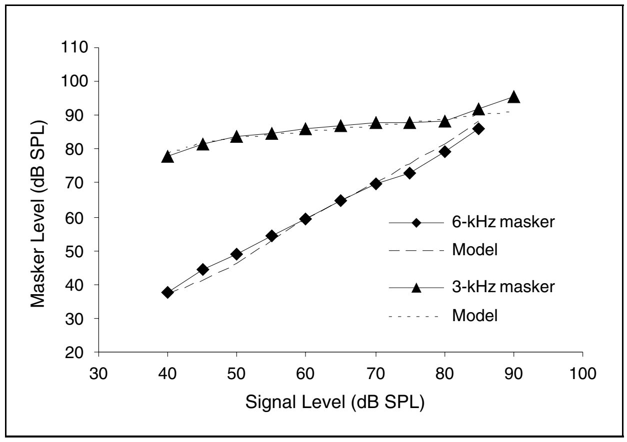

Figure 4. Forward masking data from Oxenham and Plack [34] (solid points) compared to predictions of the DRNL temporal window model (dashed lines). The signal was a 4-ms, 6-kHz pure tone preceded by a 104-ms tonal masker. The masker frequency was either 6 kHz or 3 kHz.

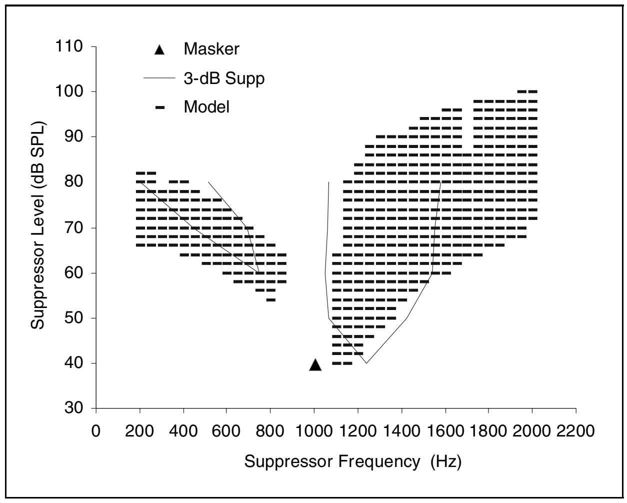

Figure 5. Suppression regions for a 1 kHz test tone from Houtgast [39] (solid lines) compared to suppression regions predicted by the DRNL filter (dashes). Regions within the solid lines, or indicated by dashes, produced at least 3 dB of suppression in Hougast’s data.

1-kHz masker tone. Those combinations of suppressor frequency and level that produced a *reduction* in the output of the model of at least 3 dB are plotted in Figure 5. (Since the 1-kHz masker tone was within the more linear portion of the BM response curve, it does not make a great deal of difference whether the suppression is measured in terms of a reduction in *excitation* level or as a reduction in the corresponding *physical* level of the stimulus[1].) The model reproduces the overall form of the suppression function

[1] A simulation was also conducted using a higher masker level of 50 dB SPL, and taking into account the effects of compression on the signal. (The signal level was varied in the experiments to find pulsation threshold.) The results were broadly similar to those presented in Figure 5. However, when expressed relative to masker level, the suppression region was a little more extensive (extending to relative suppressor levels of around 5 dB lower) both above and below the masker frequency.

quite well, although the model produces more suppression than is seen in the behavioural data for high suppressor frequencies and levels.

The model can reproduce suppression because of the combined effects of the two gammatone filters in the non-linear pathway (see Figure 2). In the suppression region, the first gammatone filter passes both the suppressor and the masker tone. These are then compressed together by the nonlinearity. The second gammatone filter in the non-linear pathway largely removes the frequency component corresponding to the suppressor. However, because the masker tone is compressed with the suppressor, its level at the output of the second filter is less than it would be if it were presented alone. The second filter in the DRNL model is intended to represent an aspect of the BM response. However, it is also possible that a filter may exist between the BM and the auditory nerve [78]. Such a filter could help to explain some of the differences between suppression on the BM and in the nerve, for example, the greater effect of below-CF suppressors in the neural data [31, 79].

If this analysis is correct then it follows that the form of the suppression contour describes the tuning properties of the first gammatone filter. It can be seen from the data that the centre frequency of the suppression region appears to be higher than the masker-tone frequency. This is reproduced in the model by giving the first gammatone a centre frequency slightly higher than CF.

Simulating suppression in this way does not require an explicit temporal window stage, since no temporal masking is involved. However, the temporal window is needed to simulate the next set of data. PTCs measured using forward masking are sharper than those measured using simultaneous masking [39, 42, 43, 44]. The difference may be the result of suppression. In forward masking the signal is masked when excitation from the masker falls within the auditory filter centred on the signal. In simultaneous masking, the signal threshold may be elevated by excitation from the masker, and by the masker suppressing (i.e. reducing the effective level of) the signal [36, 45].

The stimuli used in the simulations replicate the stimuli used in Experiment 1 in the Moore *et al.* [43] study, with the exception that notched noise (included to limit off- frequency listening) was not added to the signal and the masker. (The spectrum levels of the notched noise used in the experiment, between 5 and 18 dB, were too low compared to the masker level to have had any significant suppressive effects.) The signal frequency was 1 kHz. The masker was a 50-Hz wide band of noise with a steady-state duration of 200 ms and 10-ms ramps. The signal was a sinusoid with no steady state and 10-ms ramps. The signal was presented either simultaneously with the masker (with a 10-ms delay between signal offset and masker offset) or immediately after the masker offset (forward masking, no gap). The signal level was fixed at 52 dB SPL for simultaneous masking and at 38.5 dB SPL for forward masking. The absolute threshold for the signal, used to calculate the constant added to the output of the temporal window, was 18.5 dB SPL.

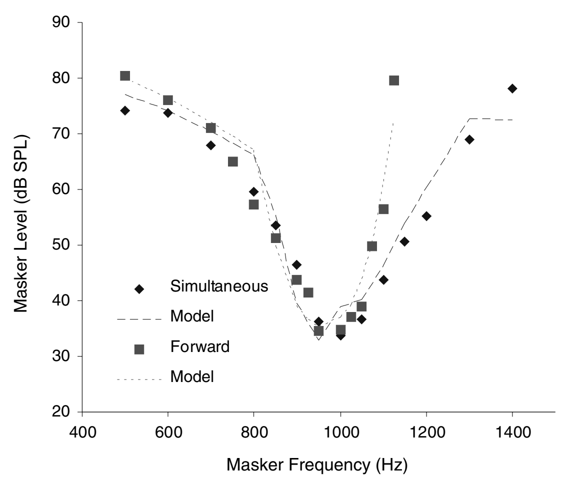

Figure 6. Psychophysical tuning curves (PTCs) from Moore *et al.* [43] (solid points) compared to predictions of the DRNL temporal window model (broken lines). The 1-kHz tonal signal was presented either simultaneously with, or following, a 50-Hz wide narrowband noise masker. Masker frequency refers to the centre frequency of the noise band. Masker level refers to the overall level of the noise band. The same model parameters were used for predicting the PTC for simultaneous masking (diamonds and dashed line) and the PTC for forward masking (squares and dotted line).

The predictions of the model were compared with the mean data from the two listeners from the Moore *et al.* study, BG and JW. To find the best fit to the data, the model parameter k (the detection efficiency) was varied adaptively. The best fitting value was 3.94 dB. This value was higher than the 1.41 dB that provided the best fit to the upward spread of masking data. It is possible that the discrepancy was the result of differences in the shape of the temporal window, in the precise form of the cochlear nonlinearity, or in detection efficiency, between the listeners in the two experiments. In forward masking, threshold differences of 10 dB or more can be observed between individuals in some conditions [65]. It is also conceivable, however, that the decision device in the model is too simplistic to account for different signal characteristics and masking paradigms.

It can be seen from Figure 6 that the model captures the main characteristics of the data quite well. The PTC for forward masking is sharper than the PTC for simultaneous masking, particularly on the high-frequency side. This reflects the greater influence of suppression for frequencies above CF, as observed in the pulsation threshold results of Houtgast. The model demonstrates that the pattern of suppression measured psychophysically (and physiologically in the auditory nerve) is roughly consistent with the differences seen between simultaneous and nonsimultaneous frequency selectivity. Again, this is consistent with the notion that the main nonlinearity in this form of masking is cochlear in origin.

It is important to understand the mechanism of masking in the level and frequency region between the PTC measured in forward masking and the PTC measured in simultaneous masking. One possibility is that the masker is suppressing the signal to a level below absolute threshold; the other possibility is that the masker is both suppressing the signal and providing excitatory masking [36]. This possibility can be investigated in the model by simply removing the constant that causes absolute threshold. If masking in the suppressive region is still observed, then there must be some excitatory masking involved. Removing the absolute threshold limit was found to have no significant effect on the form of the simulated PTCs. This means that, in the model at least, the masker is reducing the effective level of the signal to the point at which it is masked by masker excitation. This is consistent with the findings of Moore and Vickers [45] who showed that simultaneous masking for masker frequencies below the signal frequency could not be explained on the basis of suppression alone; some excitatory masking was also evident.

Houtgast [39] suggested that suppression could not be measured in simultaneous masking because suppression affects both masker and signal excitation equally at a given place on the BM, resulting in no change in the signal-to-masker ratio. In this view, suppression is a form of "distributed attenuation", acting from one part of the excitation pattern to another [40, 80]. Another possibility is that suppression by the masker is a form of "simple attenuation", causing a reduction in the whole excitation pattern produced by the signal and being equivalent to a simple reduction in the physical level of the signal [80]. If distributed attenuation is the mechanism for suppression, then "self" suppression can sharpen the excitation pattern of a single component [39]. If simple attenuation is correct then suppression only occurs *between* frequency components.

If the model presented here is an accurate reflection of the basic mechanisms of cochlear filtering, then suppression can be seen to be neither a distributed nor a simple attenuation [31]. Suppression acts from one frequency component to another, but only affects excitation in the region of the cochlea for which the frequencies of the suppressor and the suppressee both fall within the compressive region of the response (i.e., within the first gammatone filter in the DRNL model). In the model there is no self suppression. The effect of the nonlinear pathway is always to amplify, never to suppress, the response to a single tone falling within the frequency range near CF.

It should be noted that the model can reproduce some of the characteristics of two-tone suppression because the DRNL is a true nonlinear filter, not a "quasi-linear" filter. A quasi-linear filter uses a linear filter that changes its shape depending on the input level [13, 14]. Such a filter cannot reproduce the effects of suppression, although it can simulate upward spread of masking.

2.4. Limitations of the model

The model presented here extends the previous work on the nonlinear temporal window [26, 59, 65] to account for the differential effects of frequency. Although the model seems to provide a reasonable account of the data presented here, there are some aspects of temporal masking that the model cannot explain. Perhaps the most serious is the growth of suppression for suppressor frequencies above and below CF. Using a wide range of levels and frequencies Duifhuis [37] has shown quite clearly that the growth in suppression with suppressor level is greater for suppressors below the suppressee frequency than for suppressors above the suppressee frequency. The model cannot account for this effect, no matter how the parameters are adjusted, since the output of the first gammatone in the nonlinear pathway of the DRNL does not provide a frequency- dependent influence on the properties of the nonlinearity stage. As described in section 1.1, the upward spread of simultaneous masking may be influenced by suppression for low signal levels. Although this has not been explicitly tested, the model is unlikely to be able to account for the steep growth of masking in these situations.

The different growth rates for suppressor frequencies above and below CF are observed in the response of the BM [81] and the auditory nerve [82], however. Goldstein's model of the BM response [70, 71], in which low frequencies are expanded, added to the input near CF, and *then* compressed, can account for this differential growth. Clearly, some additional refinement of the DNRL model will be required to account for this aspect of the physiological and psychophysical data.

The model is not intended, at this stage in its development, to be a complete model of masking. It is used here to add weight to the argument that a simulation of the response characteristics of the BM, followed by a linear intensity integrator, may be all that is needed to simulate most temporal masking effects. It is clear that the DRNL filter does not account for all the properties of cochlear processing. However, it provides a reasonable approximation of some of these properties, and to the same extent the DRNL temporal window model can provide a reasonable account of those aspects of temporal masking that are dependent on these properties. Implemented as a complete filterbank, the DNRL model's relative simplicity and computational speed should make it an interesting alternative front-end processor, with or without the temporal window, for many applications and models of auditory processing and speech recognition.

3. Summary

In this article a possible cause of nonlinear psychophysical masking was considered, with specific reference to the extreme nonlinearities observed in temporal masking. The conclusions of the article can be expressed as follows:

1. The main nonlinearities observed in temporal masking (growth of forward masking, effect of masker duration,

nonlinear additivity, upward spread of masking, and suppression) are all consistent with the known nonlinear properties of the BM, measured directly in non-human mammals.

2. A revised version of the temporal window model, including a nonlinear auditory filter (derived originally from physiological data) and a linear intensity integrator, can account for the frequency-dependent nonlinearity of upward spread of masking and for some aspects of two-tone suppression. It had been shown previously that a similar single-channel model could account for the nonlinear masking effects relating to the BM response at CF.

3. The findings provide support for the hypothesis that, with regard to temporal masking, temporal processing subsequent to the cochlea can be modelled as a linear function of BM intensity.

Acknowledgements

The authors would like to thank Jont Allen and two anonymous reviewers for constructive comments on earlier versions of the manuscript. The second author is supported by NIH/NIDCD Grant R01 DC 03909. The third author is supported by EPSRC Grant GR/N07219. The collaboration between the first and second authors was facilitated by a Wellcome Trust Biomedical Research Collaboration Grant. The authors are indebted to Ray Meddis and Lowel O'Mard for their assistance with the DRNL model.

References

[1] B. C. J. Moore, B. R. Glasberg: Suggested formulae for calculating auditory-filter bandwidths and excitation patterns. J. Acoust. Soc. Am. **74** (1983) 750–753.

[2] R. D. Patterson: Auditory filter shape. J. Acoust. Soc. Am. **55** (1974) 802–809.

[3] T. Dau, B. Kollmeier, A. Kohlrausch: Modeling auditory processing of amplitude modulation. I. Detection and masking with narrowband carriers. J. Acoust. Soc. Am. **102** (1997) 2892–2905.

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

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

[6] G. A. Miller: Sensitivity to changes in the intensity of white noise and its relation to masking and loudness. J. Acoust. Soc. Am. **19** (1947) 609–619.

[7] W. J. McGill, J. P. Goldberg: A study of the near-miss involving weber's law and pure-tone intensity discrimination. Percept. Psychophys. **4** (1968) 105–109.

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

[9] B. C. J. Moore, D. H. Raab: Pure-tone intensity discrimination: some experiments relating to the "near-miss" to Weber's law. J. Acoust. Soc. Am. **55** (1974) 1049–1054.

[10] N. F. Viemeister: Intensity discrimination of pulsed sinusoids: the effects of filtered noise. J. Acoust. Soc. Am. **51** (1972) 1265–1269.

[11] E. Zwicker: Die elementaren Grundlagen zur Bestimmung der Informationskapazität des Gehörs. Acustica **6** (1956) 356–381.

[12] J. B. Allen, S. T. Neely: Modeling the relation between the intensity just-noticeable difference and loudness for pure tones and wideband noise. J. Acoust. Soc. Am. **102** (1997) 3628–3646.

[13] T. Irino, R. D. Patterson: A compressive gammachirp auditory filter for both physiological and psychophysical data. J. Acoust. Soc. Am. **109** (2001) 2008–2022.

[14] B. C. J. Moore, B. R. Glasberg, T. Baer: A model for the prediction of thresholds, loudness, and partial loudness. J. Aud. Eng. Soc. **45** (1997) 224–240.

[15] M. A. Ruggero, N. C. Rich, A. Recio, S. S. Narayan, L. Robles: Basilar-membrane responses to tones at the base of the chinchilla cochlea. J. Acoust. Soc. Am. **101** (1997) 2151–2163.

[16] E. Murugasu, I. J. Russell: Salicylate ototoxicity: The effects on basilar membrane displacement, cochlear microphonics, and neural responses in the basal turn of the guinea pig cochlea. Auditory Neurosci. **1** (1995) 139–150.

[17] W. S. Rhode: Observations of the vibration of the basilar membrane in squirrel monkeys using the Mössbauer technique. J. Acoust. Soc. Am. **49** (1971) 1218–1231.

[18] W. S. Rhode, N. P. Cooper: Nonlinear mechanics in the apical turn of the chinchilla cochlea in vivo. Auditory Neurosci. **3** (1996) 101–121.

[19] M. A. R. L. Robles, N. C. Rich: Basilar membrane mechanics at the base of the chinchilla cochlea. I. Input-output functions, tuning curves, and phase responses. J. Acoust. Soc. Am. **80** (1986) 1364–1374.

[20] W. Jesteadt, S. P. Bacon, J. R. Lehman: Forward masking as a function of frequency, masker level, and signal delay. J. Acoust. Soc. Am. **71** (1982) 950–962.

[21] B. C. J. Moore, B. R. Glasberg: Growth of forward masking for sinusoidal and noise maskers as a function of signal delay: Implications for suppression in noise. J. Acoust. Soc. Am. **73** (1983) 1249–1259.

[22] W. A. Munson, M. B. Gardner: Loudness patterns - a new approach. J. Acoust. Soc. Am. **22** (1950) 177–190.

[23] R. P. Carlyon: The development and decline of forward masking. Hear. Res. **32** (1988) 65–80.

[24] G. Kidd, L. L. Feth: Effects of masker duration in pure-tone forward masking. J. Acoust. Soc. Am. **72** (1982) 1384–1386.

[25] C. R. M. G. Kidd, L. L. Feth: Temporal integration of forward masking in listeners having sensorineural hearing loss. J. Acoust. Soc. Am. **75** (1984) 937–944.

[26] A. J. Oxenham, C. J. Plack: Effects of masker frequency and duration in forward masking: Further evidence for the influence of peripheral nonlinearity. Hear. Res. **150** (2000) 258–266.

[27] C. G. Cokely, L. E. Humes: Two experiments on the temporal boundaries for the nonlinear additivity of masking. J. Acoust. Soc. Am. **94** (1993) 2553–2559.

[28] A. J. Oxenham, B. C. J. Moore: Additivity of masking in normally hearing and hearing-impaired subjects. J. Acoust. Soc. Am. **98** (1995) 1921–1934.

[29] R. H. Wilson, R. Carhart: Forward and backward masking: interactions and additivity. J. Acoust. Soc. Am. **49** (1971) 1254–1263.

[30] M. J. Penner: The coding of intensity and the interaction of forward and backward masking. J. Acoust. Soc. Am. **67** (1980) 608–616.

[31] J. B. Allen: Nonlinear cochlear signal processing. – In: Physiology of the ear. A. Jahn, J. Santos-Sacchi (eds.). Singular Thomson Learning, San Diego, 2001, 393–442.

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

[33] P. G. Stelmachowicz, D. E. Lewis, L. L. Larson, W. Jesteadt: Growth of masking as a measure of response growth in hearing-impaired listeners. J. Acoust. Soc. Am. **81** (1987) 1881–1887.

[34] A. J. Oxenham, C. J. Plack: A behavioral measure of basilar-membrane nonlinearity in listeners with normal and impaired hearing. J. Acoust. Soc. Am. **101** (1997) 3666–3675.

[35] A. J. Oxenham, C. J. Plack: Suppression and the upward spread of masking. J. Acoust. Soc. Am. **104** (1998) 3500–3510.

[36] B. Delgutte: Physiological mechanisms of psychophysical masking: Observations from auditory-nerve fibers. J. Acoust. Soc. Am. **87** (1990) 791–809.

[37] H. Duifhuis: Level effects in psychophysical two-tone suppression. J. Acoust. Soc. Am. **67** (1980) 914–927.

[38] T. Houtgast: Psychophysical evidence for lateral inhibition in hearing. J. Acoust. Soc. Am. **51** (1972) 1885–1894.

[39] T. Houtgast: Lateral suppression in hearing. Free University of Amsterdam, 1974.

[40] B. C. J. Moore, B. J. O'Loughlin: The use of nonsimultaneous masking to measure frequency selectivity and suppression. – In: Frequency Selectivity in Hearing. B. C. J. Moore (ed.). Academic, London, 1986, 179–250.

[41] R. V. Shannon: Two-tone unmasking and suppression in a forward-masking situation. J. Acoust. Soc. Am. **59** (1976) 1460–1470.

[42] B. C. J. Moore: Psychophysical tuning curves measured in simultaneous and forward masking. J. Acoust. Soc. Am. **63** (1978) 524–532.

[43] B. C. J. Moore, B. R. Glasberg, B. Roberts: Refining the measurement of psychophysical tuning curves. J. Acoust. Soc. Am. **76** (1984) 1057–1066.

[44] L. L. M. Vogten: Low-level pure-tone masking: A comparison of "tuning curves" obtained with simultaneous and forward masking. J. Acoust. Soc. Am. **63** (1978) 1520–1527.

[45] B. C. J. Moore, D. A. Vickers: The role of spread of excitation and suppression in simultaneous masking. J. Acoust. Soc. Am. **102** (1997) 2284–2290.

[46] R. M. Arthur, R. R. Pfeiffer, N. Suga: Properties of 'two-tone inhibition' in primary auditory neurones. J. Physiol. **212** (1971) 593–609.

[47] P. J. Abbas, M. B. Sachs: Two-tone suppression in auditory-nerve fibers: Extension of a stimulus response relationship. J. Acoust. Soc. Am. **59** (1976) 112–122.

[48] M. A. Ruggero, N. C. Rich: Furosemide alters organ of corti mechanics: Evidence for feedback of outer hair cells upon the basilar membrane. J. Neurosci. **11** (1991) 1057–1067.

[49] A. J. Oxenham, B. C. J. Moore: Modeling the effects of peripheral nonlinearity in listeners with normal and impaired hearing. – In: Modeling sensorineural hearing loss. W. Jesteadt (ed.). Erlbaum, Hillsdale, NJ, 1997, 273–288.

[50] B. C. J. Moore, D. A. Vickers, C. J. Plack, A. J. Oxenham: Inter-relationship between different psychoacoustic measures assumed to be related to the cochlear active mechanism. J. Acoust. Soc. Am. **106** (1999) 2761–2778.

[51] D. A. Nelson, A. C. Schroder: Linearized response growth inferred from growth-of-masking slopes in ears with cochlear hearing loss. J. Acoust. Soc. Am. **101** (1997) 2186–2201.

[52] F. L. Wightman, T. McGee, M. Kramer: Factors influencing frequency selectivity in normal and hearing-impaired listeners. – In: Psychophysics and physiology of hearing. E. F. Evans, J. P. Wilson (eds.). Academic Press, London, 1977.

[53] B. C. J. Moore, B. R. Glasberg: Comparisons of frequency selectivity in simultaneous and forward masking for subjects with unilateral cochlear impairments. J. Acoust. Soc. Am. **80** (1986) 93–107.

[54] S. P. Bacon, N. F. Viemeister: Temporal modulation transfer functions in normal-hearing and hearing-impaired subjects. Audiology **24** (1985) 117–134.

[55] B. C. J. Moore, B. R. Glasberg: Gap detection with sinusoids and noise in normal, impaired and electrically stimulated ears. J. Acoust. Soc. Am. **83** (1988) 1093–1101.

[56] R. Plomp: The rate of decay of auditory sensation. J. Acoust. Soc. Am. **36** (1964) 277–282.

[57] N. F. Viemeister: Temporal modulation transfer functions based upon modulation thresholds. J. Acoust. Soc. Am. **66** (1979) 1364–1380.

[58] B. C. J. Moore, B. R. Glasberg, C. J. Plack, A. K. Biswas: The shape of the ear's temporal window. J. Acoust. Soc. Am. **83** (1988) 1102–1116.

[59] A. J. Oxenham, B. C. J. Moore: Modeling the additivity of nonsimultaneous masking. Hear. Res. **80** (1994) 105–118.

[60] C. J. Plack, B. C. J. Moore: Temporal window shape as a function of frequency and level. J. Acoust. Soc. Am. **87** (1990) 2178–2187.

[61] M. Müller, D. Robertson, G. K. Yates: Rate-versus-level functions of primary auditory nerve fibres: Evidence for square law behaviour of all fibre categories in the guinea pig. Hear. Res. **55** (1991) 50–56.

[62] M. Müller, D. Robertson: Shapes of rate-versus-level functions of primary auditory nerve fibres: Test of the basilar membrane mechanical hypothesis. Hear. Res. **57** (1991) 71–78.

[63] G. K. Yates, I. M. Winter, D. Robertson: Basilar membrane nonlinearity determines auditory nerve rate-intensity functions and cochlear dynamic range. Hear. Res. **45** (1990) 203–220.

[64] D. A. Goodman, R. L. Smith, S. C. Chamberlain: Intracellular and extracellular responses in the organ of corti of the gerbil. Hear. Res. **7** (1982) 161–179.

[65] C. J. Plack, A. J. Oxenham: Basilar-membrane nonlinearity and the growth of forward masking. J. Acoust. Soc. Am. **103** (1998) 1598–1608.

[66] J. M. Festen, R. Plomp: Relations between auditory functions in normal hearing. J. Acoust. Soc. Am. **70** (1981) 356–369.

[67] A. J. Oxenham: Forward masking: Adaptation or integration? J. Acoust. Soc. Am. **109** (2001) 732–741.

[68] R. Meddis, L. P. O'Mard, E. A. Lopez-Poveda: A computational algorithm for computing nonlinear auditory frequency selectivity. J. Acoust. Soc. Am. **109** (2001) 2852–2861.

[69] E. A. Lopez-Poveda, R. Meddis: A human nonlinear cochlear filterbank. J. Acoust. Soc. Am. **110** (2001) 3107–3118.

[70] J. L. Goldstein: Modeling rapid waveform compression on the basilar membrane as multiple-bandpass-nonlinearity filtering. Hear. Res. **49** (1990) 39–60.

[71] J. L. Goldstein: Relations among compression, suppression, and combination tones in mechanical responses of the basilar membrane: data and MBPNL model. Hear. Res. **89** (1995) 52–68.

[72] G. K. Yates: Basilar membrane nonlinearity and its influence on auditory nerve rate-intensity functions. Hear. Res. **50** (1990) 145–162.

[73] D. McFadden: The curious half octave shift: Evidence for a basalward migration of the travelling-wave envelope with increasing intensity. – In: Basic and Applied Aspects of Noise-Induced Hearing Loss. R. J. Salvi, D. Henderson, R. P. Hamernik, V. Colletti (eds.). Plenum, New York, 1986.

[74] A. L. Nuttall, D. F. Dolan: Steady-state sinusoidal velocity responses of the basilar membrane in guinea pig. J. Acoust. Soc. Am. **99** (1996) 1556–1565.

[75] W. M. Hartmann: Signals, sound, and sensation. Spring-Verlag, New York, 1997.

[76] C. J. Plack: Temporal resolution in the auditory system. University of Cambridge, 1990.

[77] ISO 226: Acoustics - normal equal-loudness contours. International Organization for Standardization, Geneva, 1987.

[78] S. S. Narayan, A. N. Temchin, A. Recio, M. A. Ruggero: Frequency tuning of basilar membrane and auditory nerve fibers in the same cochleae. Science **282** (1998) 1882–1884.

[79] N. P. Cooper: Two-tone suppression in cochlear mechanics. J. Acoust. Soc. Am. **99** (1996) 3087–3098.

[80] B. C. J. Moore: Mechanisms and frequency distribution of two-tone suppression in forward masking. J. Acoust. Soc. Am. **68** (1980) 814–824.

[81] M. A. Ruggero, L. Robles, N. C. Rich: Two-tone suppression in the basilar membrane of the cochlea: Mechanical basis of auditory-nerve rate suppression. J. Neurophysiol. **68** (1992) 1087–1099.

[82] B. Delgutte: Two-tone suppression in auditory-nerve fibers: Dependence on suppressor frequency and level. Hear. Res. **49** (1990) 225–246.