

# The Masking Property of the Auditory System: The Masking of Speech Signals

Johan J Hanekom

*Department of Electrical and Electronic Engineering  
University of Pretoria*

## ABSTRACT

*The masking property of the auditory system is well known in the context of two-tone masking. For complex (speech) signals, the effects of masking are less well known. This paper explores the masking of speech signals, by calculating which parts of the speech signal is inaudible because of masking. The theory for the masking of one tone by another is expanded, to establish an equation for the masking threshold. This masking threshold takes into account the masking of each frequency component on all other frequency components. Speech is then synthesized in which the supposedly inaudible parts of the speech signal are discarded, and the effects are evaluated in a very simple psychoacoustic experiment. It is shown that the information below the masking threshold is indeed redundant.*

## OPSOMMING

*Die maskeringseienskap van die gehoorstelsel is welbekend in die konteks van twee-toon-maskering. Die effek wat maskering het op komplekse (spraak) seine, is minder bekend. Hierdie artikel verken die maskering van spraakseine deur te bereken watter gedeeltes van die spraaksein is onhoorbaar. Die teorie van die maskering van een toon deur 'n ander word uitgebrei om 'n vergelyking vir die maskerdrempel te bepaal. Hierdie maskerdrempel neem die maskerings effek van elke frekwensiekomponent op elke ander frekwensiekomponent in ag. Spraak word dan gesintetiseer, met die dele wat as onhoorbaar geag word, verwyder. Die effek hiervan word in 'n eenvoudige eksperiment evalueer. Dit word aangetoon dat inligting onder die maskerdrempel wel oortollig is.*

**KEY WORDS:** auditory processing, masking functions, masking threshold, speech spectrum, spectrum synthesis, two-tone masking

## INTRODUCTION

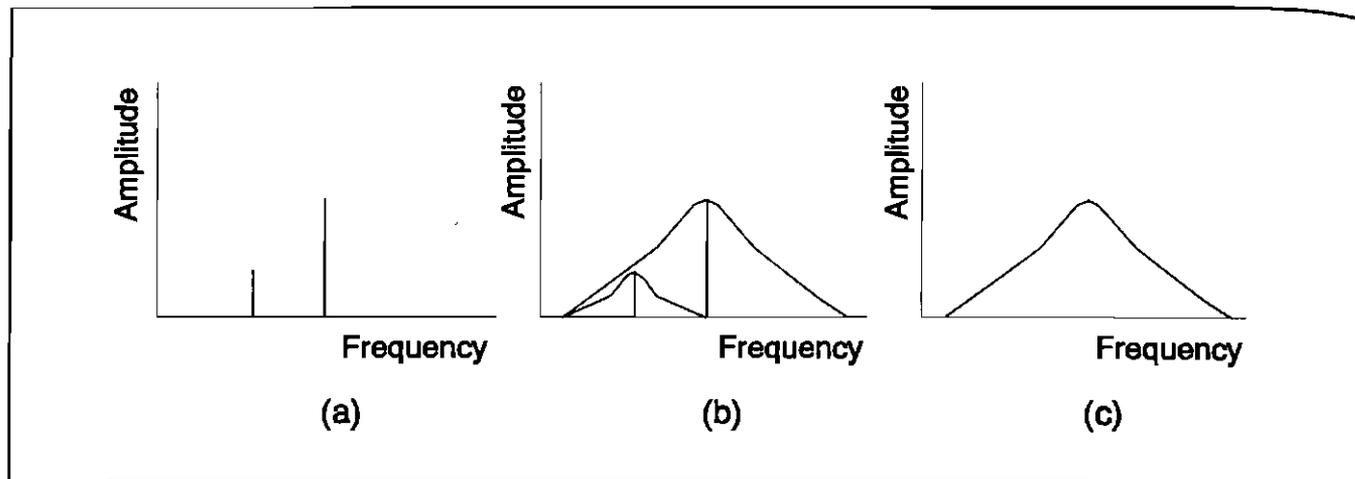
Many questions concerning masking<sup>1</sup> in the auditory system remain unanswered. On the one hand, the phenomenon of two-tone masking is well known (Javel, 1981), (Kanis & De Boer, 1994), as is the masking of a noise band by a tone, or vice versa. On the other hand, little information is available on masking and its effects in complex sounds (speech sounds). We might ask whether the masking mechanism does in fact function for complex sounds. If masking does function for complex sounds, what is the mechanism and why does the auditory system suppress some information? What are the effects of the masking? The purpose of this paper is to explore some of these questions about masking.

The approach used in this paper regards masking from a different perspective than that normally found in the literature on the subject. Instead of setting up a psychoacoustic experiment, and using various tones or complex signals to determine the masking of one signal

by another, a speech synthesis approach is used. A speech signal, in which all the parts of the signal which are supposedly inaudible owing to masking, are discarded, is synthesized.

If one tone can mask a second tone, then this second tone might also have a masking effect on a third tone (as well as on the first). The hypothesis is that each component in the speech frequency spectrum has a masking effect, however limited, on every other component of the speech spectrum. This statement implies that some parts of the speech spectrum are never heard or are *redundant*; but which parts? Can we discard these parts of the spectrum without loss of fidelity? To attempt to answer these questions, we have to develop a model which describes how each component of the spectrum masks every other component of the spectrum, almost as if each spectral component was in a two-tone contest with each other spectral component. We will use the well-known data on masking to determine mathematical expressions for *masking functions* for each spectral component of a complex signal.

<sup>1</sup> The auditory system has the characteristic that weaker spectral components are masked by stronger spectral components. This simply means that the weaker component is inaudible in the presence of the stronger component (Allen, 1985).



**Figure 1. Masking explained conceptually.** The frequency spectrum for two pure tones is shown in (a). (b) shows the activation area of each tone. (c) shows the resulting activation area: the activation area of the softer tone is swamped by the louder tone. The louder tone masks the smaller tone, and only the larger frequency component (the louder tone) is audible. The amplitude axis might represent the displacement of the basilar membrane or, alternatively, the neural firing rate.

### THE ORIGIN OF MASKING

Before we start the mathematical analysis to describe masking in the auditory system, a brief explanation of the possible origins of masking will shed light on the above statement that each spectral component masks every other spectral component.

The pathway that the auditory signal follows through the auditory system is conceptually summarised in the statements to follow. This description ignores some of the complexities of the auditory system.

A pressure wave (the sound wave) is transmitted in the air. The pressure wave is received by the antenna (the pinna) and transmitted through the outer ear canal. Tympanic membrane to cochlea transmission takes place via the middle ear structures. The pressure wave travels from the cochlear oval window down the basilar membrane. A frequency to position transformation takes place on the basilar membrane (the cochlear partition acts as a dispersive filter) (Allen, 1985). Hair cells sense the basilar membrane displacement, and in turn displacement of hair cells stimulates the generation of action potentials on the nerve endings synapsing with the hair cells (or, in other phraseology, the neurons fire). The cochlear nerve carries information in the form of action potentials, via various auditory centres, to the auditory cortex of the brain (Keidel, Kallert & Korth, 1983).

Masking is observed at several locations in the auditory system: at the hair cell level, in the coding of the neural firing patterns, as well as on the basilar membrane (Javel, 1981). The origin of masking is explained below. The description is conceptual, and for the greater part an unconfirmed hypotheses; see also Zwicker & Zwicker (1991).

The coding of *frequency* information in the auditory system adheres, according to most of the currently accepted models (Neely & Kim, 1986, and Allen, 1985), to the place theory, i.e., frequency is coded mainly by the *place of maximum activity* on the basilar membrane (a frequency to position transformation transpires). Temporal mechanisms for frequency coding also exist, e.g., phaselocking (Allen, 1985).

The coding of *intensity* information for a single tone is done according to a principle sometimes known as the volley principle (Keidel, 1980):

- the louder the sound, the wider the area of nerve activity in the vicinity of the specific frequency component's characteristic basilar membrane position; and
- the louder the sound, the higher the frequency of firing of the nerves that emanate from that specific position of maximum basilar membrane displacement.

If the activation area (the area of basilar membrane displacement as well as the nerve activation area) in the vicinity of a loud tone becomes wider, and a softer tone has a frequency near to the louder tone, then overlap might occur between the activation areas of the two tones (figures 1 and 2). The brain's auditory processor might interpret basilar activation at the softer tone's characteristic position on the basilar membrane, as if the louder tone had activated a wider area, which extends to the softer tone's basilar position. If the auditory processor ignores the softer tone, it is said to fall below the *masking threshold* of the louder tone.

If the description above is indeed a true account of the mechanism of masking, it is expected that an inaudibly soft tone (below the masking threshold) in the frequency vicinity of a louder tone, would make the louder tone sound even louder. This has been observed in cochlear implant experiments (Hanekom, 1990 and Eddington et al., 1978), and the mechanism is called *sensitizing*.

In summary, this explanation simply means that a softer tone is swamped by a louder tone, and that in this swamped condition the specific neural channel normally used by the softer tone is unavailable.

Two further explanations for the masking phenomenon exist. In addition to the foregoing explanation, masking is also explained by the inability of the hair cells to have a displacement much greater than the displacement already caused by the louder tone, so that the softer tone has little additional effect on hair cell displacement.

Lateral inhibition between adjacent neural pathways is an additional potential contributing factor in masking.

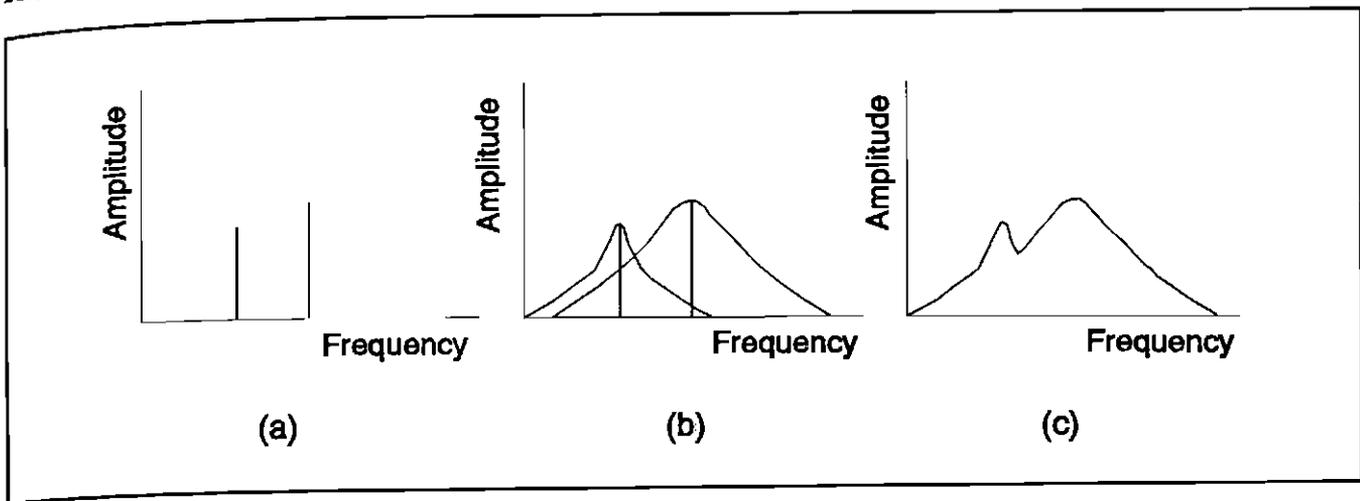


Figure 2. Masking in the case of two strong frequency components (a). (b) shows the activation area for each tone. In this case, the weaker component still influences the resulting activation, and thus both tones are audible (c). No masking takes place.

This simply means that a high frequency of neural activity on a specific neuron can suppress activity on adjacent neurons.

From the explanation above, it is clear that the further the softer tone is away from the louder tone on the spectral plane, the less the masking influence of the louder tone on the softer tone. Many examples of studies of two-tone suppression can be found in the literature (see for example Terhardt, 1979, Javel, 1981 and Kanis & De Boer, 1994), and also of the suppression of a tone by bandlimited noise, or vice versa.

Terhardt (1979) made an analysis of the processes involved in masking, and fitted models to available data from the literature, in order to design mathematical equations for the characterization of masking. The experiments on which this paper reports, apply the theory developed by Terhardt (1979). The theory, which is briefly elucidated below, refers to these equations as the *masking functions*, because they define a masking threshold in the frequency domain. The sections of the sound signal below this threshold are supposedly inaudible.

**MATHEMATICAL ANALYSIS: THE MASKING FUNCTIONS**

In the explanation to follow, the data from two-tone experiments are extended to an equation that gives the sum of the masking effects of all the frequency components in the speech spectrum, on a specific tone somewhere in the spectrum. Thus, for each spectral position, a masking threshold is calculated. If this masking threshold is known for each spectral position, the *masking threshold function* for the spectrum in its entirety is known. This masking threshold function can be found as an explicit equation, as will be shown below. According to the hypothesis, all spectral components with amplitudes below this masking threshold are inaudible and are regarded as redundant. We should be able to discard this information from the spectrum with no loss in fidelity. We will test the truth of this statement in an exploratory psychoacoustic experiment, which is described following the mathematical analysis.

As the first step in finding the masking function of a specific single tone, the frequency of the tone (in Hz) is

transformed to the *Bark scale*. The symbol for frequency on the Bark scale is  $z$  and on this scale, frequency is known as the *critical band rate*. The motivation for the use of the Bark scale will be clarified below.

The equation for the translation of frequency to Bark is given in terms of the *arctan* function (Terhardt, 1979):

$$z = 13.3 \arctan(0.75 f) \text{ Bark} \tag{1}$$

(where  $f$  is the frequency in kHz) or alternatively, in terms of *hyperbolic sine* (Schroeder, Atal & Hall, 1979):

$$f = 650 \sinh(z/7) \tag{2}$$

These equations have been determined empirically by the authors, to fit measured (psychoacoustic) data. As an example, the transformation equation (either equation (1) or equation (2)), when applied, translates 0 Hz to 0 Bark and 4 kHz to 16.6 Bark. The frequency interval from 0-1 Bark (0-100 Hz) is known as the *first critical band*, with the *second critical band* from 1 to 2 Bark (100 Hz to 310 Hz). These critical bands increase in width with higher frequency, which means that the masking functions, which are functions of the critical band rate  $z$ , become wider for higher frequency tones. This in turn means that the frequency resolution of hearing decreases at higher frequencies. The Bark scale is convenient, in that the masking functions are linear on this scale, and all masking functions throughout the spectrum have the same shape and width, whereas on a linear frequency scale, the masking functions become wider at high frequencies. This explains why the Bark scale is sometimes preferred in descriptions of auditory function.

The masking functions can now be calculated. The amount of masking by a tone of frequencies lower than itself, is found to be  $S_1 = 27 \text{ decibel(dB)/Bark}$  (Terhardt, 1979) and masking of frequencies higher than the masker tone is dependent on the specific *sound pressure level* (SPL) value of the masker tone, as well as the frequency of the masker tone, and is given by

$$S_2 = [24 + 0.23 (f_p)^{-1} - 0.2 L_p] \tag{3}$$

in dB/Bark. This is the equation for a straight line: the slope of the masking functions are linear on the Bark scale.  $S_2$  is the slope towards the higher frequencies,  $S_1$  is the slope towards the lower frequencies,  $f_v$  is the frequency in kHz of the masker tone, and  $L_v$  is the level (in dB SPL) of the masker tone.

Next, we determine how much of the softer tone, which is being masked (the maskee), protrudes above the masking threshold. The value for the masker threshold at  $f_\mu$  is simply the equation for a straight line:

$$\begin{aligned} L_{\mu v}^{(1)} &= L_v - S_2(z_\mu - z_v) \text{ for } f_v < f_\mu \\ L_{\mu v}^{(2)} &= L_v - S_1(z_v - z_\mu) \text{ for } f_v > f_\mu \end{aligned} \quad (4)$$

$f_\mu$  is the frequency of the maskee.  $z_\mu$  and  $z_v$  are the frequencies of the maskee and the masker on the Bark scale, respectively.  $L_{\mu v}^{(1)}$  and  $L_{\mu v}^{(2)}$  are the amounts by which the maskee values exceed the masking thresholds, for a maskee to the right and to the left of the masker tone, respectively.

If the masking threshold is not exceeded by the maskee, the maskee is inaudible. Thus, theoretically, the inaudible parts of the spectrum can be removed without a listener being able to perceive the difference.

The masking function as depicted above, describes how one tone masks another tone. It seems intuitively obvious that to find the masking thresholds that operate on a specific frequency component, as a result of all the other frequency components, the preceding theory could be expanded to establish the sum of the effects of all the masking tones. If we want to determine the masking effect of each frequency component in the spectrum on every other frequency component, this sum can be derived from equation (4):

$$L_v = 20 \log A_v, \text{ with } A_v = \sum_{\mu=1}^{\mu-1} 10^{L_{\mu v}^{(1)}/20} + \sum_{\mu=\mu+1}^N 10^{L_{\mu v}^{(2)}/20} \quad (5)$$

This equation calculates a value for the masking threshold. Note that the sound pressure amplitudes in Pascal/m<sup>2</sup> are summed, and not the dB SPL values. This sum is then converted back to dB SPL.

The two summations are used to calculate the contributions to the masking of respectively all the components lower, and all the components higher than the specific maskee frequency under consideration ( $f_v$ ). For frequency components higher than the maskee frequency  $f_v$ , masker contributions are calculated by taking into account their masking threshold slopes on their lower frequency sides ( $S_1 = -27$  dB/Bark). For frequencies lower than  $f_v$ ,  $S_2$  from equation (3) is used.

This analysis is adequate for exploratory experiments on the effects of masking in speech.

## METHOD

For a two-tone experiment, masking is easily established in a psychoacoustic experiment (Javel, 1981). In order to investigate in a psychoacoustic experiment whether masking does occur in the auditory processing of the complex speech spectrum in the way predicted by equation (5), the test will be whether or not the information theorized to be redundant (the information below the masking threshold calculated from equation (5)), is audible or inaudible. As a first exploration, a simple psychoacoustic experiment was devised.

The equations above (1-5) were implemented in a computer program. The program takes normal speech as input, and outputs a "distorted" version of this speech signal (all information below the masking threshold is regarded as redundant and is discarded). The operation of the program is briefly described.

The input signal is a file of prerecorded speech data. The data comes from a calibrated microphone, and as such each value of the data is a digital representation of a voltage. Data samples were taken at a frequency of 8 kHz. The voltage values can be converted to SPL values if the characteristics of the microphone are known. For the conversion the equation used is

$$V(\mu V) = 10^{0.0475 \text{ SPL(dB)} - 0.7}$$

which was established empirically for the specific microphone used.

After the conversion to SPL values, the time domain signal is transformed to the frequency domain using the Fast Fourier Transform. The masking threshold in the frequency domain is then calculated according to the equations given earlier (5). The masking threshold is then compared to the spectrum of the original signal, and where the spectrum does not exceed the threshold, the spectral information is discarded. Discarding of sections of the spectrum does not mean that we can merely make those values zero, because zeros in the spectrum cause echoes in the resultant sound. A *discarding function* was therefore implemented, as explained later. A minimum of 10 dB was chosen as the minimum value that any spectral component can assume. 10 dB was chosen as a minimum, because it is far below the normal 30-40 dB ambient noise. After thresholding, the thresholded spectrum is transformed to the time domain by the Inverse Fourier Transform. This data is then output through a digital to analog converter, amplifier and loudspeaker.

The quality of speech after masking could only be determined qualitatively because of a lack of quantitative measures of speech quality. Mathematical measures, e.g., Mean Square Error, is inadequate for the measurement of speech quality. A reasonable objective measure is described in Schroeder (1979), where the masking functions are used to calculate a single value as a measure for quality.

For reliable qualitative determination of sound quality, a reference is needed, and this reference is used in paired comparison tests. Two references were used. The original signal was used as the one reference, and the other reference was the speech signal thresholded by a level threshold, which was initially set at 25 dB SPL. 25 dB was used as threshold, as with this choice about 50 % of the signal spectrum was below the threshold, which was more or less the same amount of data below the calculated masking threshold for the specific input speech signal. The speech signal, distorted by applying the calculated masking threshold and discarding redundant information, was then compared to these two reference signals. In further experiments the threshold was translated linearly upward, resulting in more of the original spectrum falling below the threshold and therefore being discarded. The purpose is explained in the discussion. The same linear translation was done with the level threshold, always ensuring that the percentage of discarded data remained similar for the level threshold and the threshold calculated from the masking functions.

The purpose of this experiment was to establish whether the discarding of supposedly redundant information was perceptible. Uninformed listeners were asked to grade the quality of three different speech signals: the original, the signal distorted by a level threshold, and distortion by a threshold calculated from the masking functions.

Two implementations of the discarding function were used: (1) the discarded values were set equal to 10 dB, (2) the discarded values were taken as  $value(n) = value(n-1) \times 0.9$ . This simply gives a gentle decay to 10 dB, instead of an abrupt transition. Deep holes in the spectrum have the perceptual effect of sounding like echoes. Also, normally abrupt transitions carry speech information (e.g., the sharp transitions found in start and stop consonants). Thus, the way in which the redundant data are discarded, influences the perceptual quality of the thresholded speech signal, while not having any relation to the effects of masking.

To establish the occurrence of masking in the way predicted by equation (5), we simply need to demonstrate that random alterations can be made to the part of the signal below threshold, without any perceptible difference in the signal. Any alteration is fine, on two conditions: (1) no deep holes in the spectrum are allowed and (2) the changed section of the signal must still be below threshold.

**RESULTS**

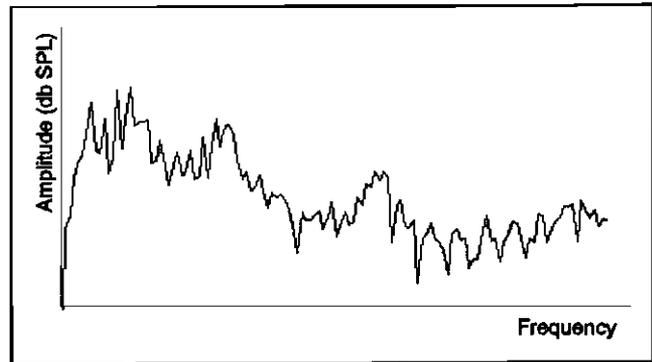
Examples of the thresholding process and the resultant signal are given in figures 3 and 4.

The results of the grading experiments are given in table 1 below. *Method 1* refers to the method in which discarded values are set equal to 10 dB. *Method 2* refers to the method in which the gentle decay function was implemented. The percentages refer to the amount of data that has been discarded. The discarding of approximately 50 % of the original spectral data occurs for the specific input speech (phonetically balanced sentences) when equation (5) is applied. Thus, the 50 % case in the table is without any upward translation of the threshold curve. The numbers in the table refer to the grading given by the listeners, where 1 is the best and 4 the worst. Where the same grading is given in two columns, the differences between these two sounds were imperceptible.

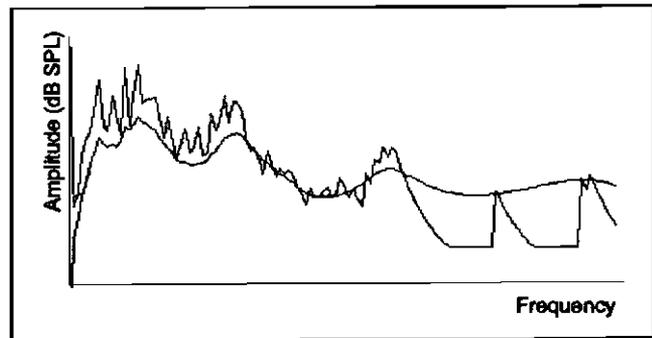
With 50 % of the signal below threshold, no difference between any of the signals is discernible. Although this might seem amazing, most of the data that were discarded, were at the higher frequencies (figure 4), where the frequency sensitivity is not as high. This means that the periodic time structure of the time domain signal is well-preserved, and no audible pitch distortion is observed.

At 75 % discarded information, the difference between the threshold signals and the original becomes audible, although not considerably. Interestingly, the quality of

sound from the level threshold was rated the same as the masking threshold. At 90 %, method 2 gave the best thresholded sound quality. The level threshold gave the worst sound quality by far. In both the 75 % and the 90 % case, the method 2 sounded better than method 1.



**Figure 3.** The original spectrum before thresholding, plotted from 0 Hz to 4000 Hz (x-axis). The y-axis gives the spectral amplitude in dB SPL. The scaling is not shown and is not important.



**Figure 4.** The spectrum after thresholding, plotted from 0 Hz to 4000 Hz (x-axis). The y-axis is the amplitude in dB SPL. The threshold is the smooth line. The jagged line is the spectrum after thresholding. The part of the spectrum above the threshold is retained. The part below the threshold is the spectrum after application of the discarding function. The effect of the gentle decay discarding function can be seen clearly at the high frequency side of the spectrum.

**DISCUSSION**

As is evident from the results, masking does seem to occur in the auditory processing of complex (speech) signals in the way predicted by equation (5). For the specific speech signal used in this simple experiment, alterations in the supposedly redundant sections of the spectrum were

**Table 1. Results of the grading of the speech signal quality, before and after masking**

Threshold set at:	original	method 2 level threshold	method 1 masking threshold	method 2 masking threshold
50 %	1	1	1	1
75 %	1	2	3	2
90 %	1	4	3	2

inaudible. Although it is not the only information-reduction process in the auditory system, masking does play an important information-reduction role. With masking function based distortion, even with 90 % of the original signal discarded, the speech is still easily comprehensible, although the speech quality has decreased. Masking eliminates some of the redundancy in the signal. Using the original calculated masking threshold (without translation), the information rate is cut by about 50 %, without any audible reduction in sound quality.

The purpose of the linear translation of the masking threshold was to explore the possibilities of using the calculated masking threshold for engineering applications. This shifted threshold is artificial and does not have any direct significance in a description of the functioning of masking in the auditory system. The information being discarded is not redundant and audible distortion is expected. Distortion is, however, applied in a controlled way, and we are not discarding more important information from some sections of the spectrum than from other sections, as we are doing when a level threshold is applied.

Engineering applications of the masking thresholds as they are described here, are among others in speech coding. With a preprocessor based on the masking thresholds of the normal ear, one can apply controlled distortion onto a speech signal to reduce the information rate.

As explained earlier, the fact that approximately 50 % of the spectrum was discarded with the calculated masking threshold, was used to determine the level for the level threshold. Although the difference between the level threshold and the threshold determined by the masking functions is not directly evident in the 50 % experiment, from the sound quality observed in the 75 % and 90 % experiments it is conceivable that the calculated masking functions approximate the masking process in the auditory system.

Improvements could be made to the model used for masking in this paper, e.g., by basing the model not on psychoacoustic experimental data, but on physiological data. As has been explained, results from two-tone masking experiments have been used to determine the masking functions which were used in these experiments. The two-tone masking functions were expanded in equation (5) for application to more complex spectra. Possibly, this expansion is not the most applicable masking model to implement on complex speech spectra, as was done here. However, no measured data on the masking observed in complex spectra are available (although data for tone/handlimited noise masking are available). This might account for the somewhat strange result, that the signal distorted by the level threshold sounded almost the same as the signal distorted by masking threshold (in the 75 % case).

The discarding function is not based on any measured data. It is not possible to determine from psychoacoustic experiments how the data reduction in masking is implemented into neural firing patterns. From the description in the introduction, it can be guessed that the information is not suppressed, as in the implementation, but simply swamped. That masking operates like a swamping (or saturation) function and not an attenuation function, is

motivated by Kanis & De Boer (1994) and Javel (1981).

The discarding function was implemented here in order to demonstrate that the frequency components below the threshold are inaudible, and not to try to simulate the normal auditory functioning. Actually, we might just as well have distorted the sections of the signal below threshold in any other way to prove that these distortions would be inaudible. When we do this, we have to comply with at least the two rules stated earlier, and a third rule may be implemented with flexibility:

- The amplitude in the sections of the signal that are to be distorted must stay within the same bounds as the amplitude of the original signal in these regions.

## CONCLUSION

Masking plays an important role in the data-reduction mechanism of the peripheral parts of the auditory system. Although the experiment described here was meant to be exploratory rather than conclusive, the result indicates that the understanding of the mechanism of masking that led to equation (5), seems to be reasonable. In order to gain a better understanding of the complexities of auditory processing, it is important that the masking property of the auditory system is not studied in isolation from the other characteristics of auditory processing. On the one hand, the psychoacoustic study of the masking of complex signals should be expanded. On the other hand, more cohesive cochlear models, based on the physiology rather than being heuristic, should be created to assimilate the available data.

## REFERENCES

- Allen, J.B. (1985). Cochlear Modelling. *IEEE Acoustics, Speech and Signal Processing Magazine*, 2, 1, 3-29.
- Eddington, D.K., Dobelle W.H., Brackmann, D.E., Mladejovsky, M.G. & Parkin, J.L. (1978). Auditory Prostheses Research with Multiple Channel Intracochlear Stimulation in Man. *The Annals of Otolaryngology, Rhinology & Laryngology*, 53, 87, 6, 1-39.
- Hanekom, J.J. (1990). *Die Ontwikkeling van 'n Suid-Afrikaanse Bioniese-Oor*. Master's Dissertation. Department of Electrical and Electronic Engineering: University of Pretoria.
- Javel, E. (1981). Suppression of Auditory Nerve Responses I: Temporal Analysis, Intensity Effects and Suppression Contours. *Journal of the Acoustical Society of America*, 69, 6, 1735-1745.
- Kanis, L-J. & de Boer, E. (1994). Two-tone Suppression in a Locally Active Nonlinear Model of the Cochlea. *Journal of the Acoustical Society of America*, 96, 4, 2156-2165.
- Keidel, W.D. (1980). Neurophysiological Requirements for Implanted Cochlear Prosthesis. *Audiology*, 19, 105-127.
- Keidel, W.D., Kallert, S. & Korth, M. (1983). *The Physiological Basis of Hearing: A Review*. New York: Thieme-Stratton Incorporated.
- Neely, S.T. & Kim, D.O. (1986). A model for active elements in cochlear biomechanics. *Journal of the Acoustical Society of America*, 79, 5, 1472-1480.
- Schroeder, M., Atal, B.S. & Hall, J. L. (1979). Optimizing Digital Speech Coders by Exploiting Masking Properties of the Human Ear. *Journal of the Acoustical Society of America*, 66, 12, 1647-1652.
- Terhardt, E. (1979). Calculating Virtual Pitch. *Hearing Research*, 1, 155-182.
- Zwicker, E. & Zwicker, U.T. (1991). *Audio Engineering and*